Big Data Approaches to Improving the Identification of Drug or Disease Mechanisms for Drug Innovation

by

Muying Wang

Master of Science, University of Pittsburgh, 2017

Submitted to the Graduate Faculty of

the Swanson School of Engineering in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2020

UNIVERSITY OF PITTSBURGH

SWANSON SCHOOL OF ENGINEERING

This dissertation was presented

by

Muying Wang

It was defended on

February 27, 2020

and approved by

Jason E. Shoemaker, Ph.D., Assistant Professor, Department of Chemical and Petroleum Engineering

Ipsita Banerjee, Ph.D., Associate Professor, Department of Chemical and Petroleum Engineering

James R. Faeder, Ph.D., Associate Professor, Department of Computational and Systems Biology

Christopher E. Wilmer, Ph.D., Assistant Professor, Department of Chemical and Petroleum Engineering

Dissertation Director: Jason E. Shoemaker, Ph.D., Assistant Professor, Department of Chemical and Petroleum Engineering

Big Data Approaches to Improving the Identification of Drug or Disease Mechanisms for Drug Innovation

Muying Wang, PhD

University of Pittsburgh, 2020

Advances in science and technology have substantially changed drug research and development (R&D) processes. However, the efficiency of drug R&D, described in the number of new drugs approved per billion US dollars spent, dramatically declined between 1950 to 2010. Some of the main causes to the attrition include the cautious regulator, the potential risk in chemical screening methods for early drug discovery, and the lack of understanding of disease mechanisms. In order to improve the efficiency and productivity in drug R&D, more powerful tools are needed to assist in prediction forecasting and decision making in drug development.

This dissertation describes my work in developing computational approaches to provide better understanding of drug or disease mechanisms at the systems level. The first project involves collaboration with RIKEN institute in Japan for innovation of influenza vaccine adjuvant. We performed comparative analysis of RNA-Seq data from mice treated with different adjuvants to identify mechanisms supporting adjuvant activity. In the second project, we predicted immune cell dynamics by linear regression-based algorithms or statistical tools, and suggested a new approach that can improve the discovery of key disease-associated genes. In the third project, we found that the network topological features, especially network betweenness, predominantly define the accuracy of a major drug target inference algorithm. We proposed a novel algorithm, TREAP, which integrated betweenness and differential gene expression and can accurately predict drug targets in a time-efficient manner. Through the projects above, we have demonstrated how computational algorithms can assist in mining big biological data to improve understanding of drug or disease mechanisms for drug innovation and development.

Table of Contents

List of Tables

List of Figures

Preface

This work and my graduate program started with the interests in traditional Chinese medicine and pharmacy. As a teenager, I read several books and articles about a Chinese physician, Zhongjing Zhang. Living around 1800 years ago, he not only summarized and established the principles of Chinese medicine but was also revered for treating the disadvantaged, regardless of their social status. I have been inspired by the advanced ideas in his works, such as regarding the human body as a system, designing personal treatments, controlling toxicity of a prescription, and utilizing one prescription for treating multiple diseases. For example, eyes are believed to be a dependent to the liver. Therefore, any symptoms of eyes are reflection of disorders in the liver, and to treat the eyes, one should diagnose and react to the actual problem with the liver. What's more, with sympathy for patients, Zhang's drug prescriptions were designed as inexpensive as possible while maximizing their effects. With him as an example, I intend to work on life science or healthcare research and innovation that can increase efficiency of drug treatments, improve patients' experience, and lower the cost for society.

Living in 21st century, we can achieve better and more efficient diagnosis and treatment. However, people are still facing challenges in some complex diseases, such as cancer, whose mechanisms or treatments have not yet been uncovered. Some of the barriers lie in the lack of principles similar to Chinese medicine which can explain human body and disease pathology at the systems level.

When I was in high school, I learned about technologies related to big data and bioinformatics approaches. I found it promising to use these approaches to understand disease mechanisms and innovate new drugs, because they have potential in capturing characteristics of biological samples at various levels, from a single cell to a live animal. Understanding the changes that computational tools and artificial intelligence will make to the life science and healthcare industry, I decided to learn programming during my graduate program and it is my motivation for this work.

This work would not have been possible without my advisor, Dr Jason Shoemaker. He has always been supportive for my research projects. And I am sincerely grateful for all of his help with my graduate education and suggestions on career paths. In addition, I would also like to thank all of my committee members for their time and advice on my proposal and this work. And I would like to thank our collaborator, Dr. Hidehiro Fukuyama at RIKEN, for the research opportunities.

I would also like to thank my colleagues, Shoemaker Lab members, Emily Ackerman, Robert Gregg, Jordan Weaver, Lauren Luciani and Ericka Mochan. I am grateful for and deeply influenced by the positive and warm environment they have all created. I would like to attribute special thanks to Siying Zhang. She has always been there for me as a friend, colleague and sometimes mentor to help me overcome any difficulties. Finally, I would like to thank my parents for encouraging and supporting my education.

This work includes three peer-reviewed articles, which are published in *IFAC* [1], *Processes* [2] and *International Immunology* [3]. One more article is under review at the time of writing [4], and another article by our collaborators is expected. The source codes of TREAP (the R package) are available for public use on GitHub.

1.0 Background

1.1 Challenges of Drug Research and Development

Advances in science and technologies since 1950s have substantially changed drug research and development (R&D) processes. For example, high-throughput screening can conduct thousands of tests simultaneously, and it has been broadly utilized in pharmaceutical industry and contributed to the discovery of drugs, such as chemokine receptor antagonists [5]. Other breakthroughs associated with pharmaceutical industry include but are not limited to DNA sequencing and X-ray crystallography. However, the efficiency of drug R&D, described in the number of new drugs approved per billion US dollars spent (inflation adjusted), dramatically declined between 1950 to 2010 [6].

The main causes to the attrition in efficiency and existing challenges in drug R&D lie in (1) the cautious regulator, (2) the potential risk in chemical screening methods for early drug discovery, and (3) the lack of understanding of disease mechanisms [6], which will be further explained in the following sections.

1.1.1 The Cautious Regulator and Clinical Trial Failures

The cautious regulator has shown obvious impact on the efficiency of drug R&D. After the tragedy of thalidomide, the U.S. Food and Drug Administration (FDA) tightened the regulation on new drugs, so that the efficiency of drug R&D decreased from 1960s to 1970s [6]. On the contrary, in 1990s, the efficiency temporarily increased as the FDA cleared backlog and lowered the regulatory barriers for HIV drugs [6].

The regulatory processes for the introduction of new drugs have been profoundly changed since 1950s. Currently, a typical drug development pipeline contains 5 stages as described by the FDA: (1) discovery and development, (2) preclinical research, (3) clinical research, (4) FDA drug review and (5) FDA post-market drug safety monitoring [7] (Figure 1).

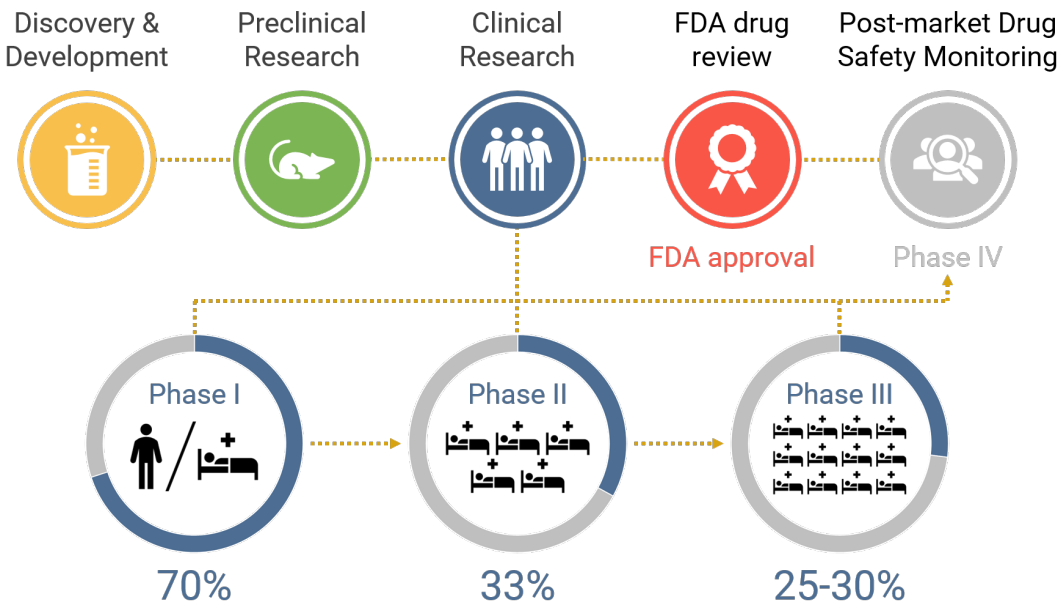


Figure 1: Overview of the drug development pipeline and clinical trial phases.

At the stage of discovery and development, new drug candidates are often discovered through improved understanding of disease pathology or tests of molecular compounds, such as high-throughput screening. Experiments are performed to understand the efficacy, mechanisms of action, dosage and toxicity of promising drug candidates [8]. The information gathered help to decide which candidates will proceed to preclinical and clinical researches. For preclinical research, the dosing and safety of drug candidates are tested *in vitro* or *in vivo* before they are tested in people [9]. The findings assist in decision-making for clinical trials.

The clinical research, which contains four phases, is designed to study the efficacy, toxicity and dosage of drug candidates by testing them in people [10] (Figure 1). In Phase I, drug candidates are tested in around 20 to 100 healthy volunteers or people with the disease/condition to understand their safety and dosage. According to FDA, about 70% of all candidates will move to Phase II [10]. In Phase II, several hundred people with the disease/condition are tested for a few months to two years to collect data about the efficacy

and side effects of drug candidates. At this phase in clinical trials, merely 33% of candidates can move to Phase III [10]. Phase III is the last step before FDA review and approval. The studies are larger in scale, with 300 to 3000 people with the disease/condition, and longer in duration, one to four years for the length of studies. This phase provides most of the data on the safety and the benefits that a drug candidate can offer to a specific population. Only 25-30% of drug candidates can move forward to FDA review. Usually after sufficient data is collected from Phase I to Phase III trials, a New Drug Application is filed and the FDA will make a decision on whether to approve this drug. Once it is approved, the drug developer will be able to market the drug, and Phase IV trials are carried out to monitor its safety [11].

Based on the success rates of each phase, the overall rate for a drug candidate to pass through Phase I to III and be approved is around 7%. It is necessary to carefully study the reasons for failures in clinical research and optimize the drug development pipeline to improve the efficiency accordingly [6, 12]. From 2008 to 2015, the primary cause to failure in Phase II and Phase III maintained to be insufficient efficacy, which accounted for around 50-65% of all failures in each phase [12, 13, 14]. This is surprising because early drug discovery and preclinical tests are supposed to diminish the possibility of failures due to efficacy. What's more, the percentage of insufficient efficacy at Phase III was comparable with Phase II. The secondary cause was safety issues, and the percentage was around 15-35% in each phase. Combining the primary and secondary causes, the clinical trial failures caused by insufficient efficacy or safety issues accounted for 70-85% of all failures at Phase II or III. The rest of the causes are commercial, strategic and operational concerns or issues. The above suggests that currently procedures in terms of early drug discovery and preclinical tests may need improvements to reduce failures in clinical research.

1.1.2 The Potential Risk in Chemical Screening Methods for Early Drug Discovery

New drugs approved by the FDA in the past few years can be roughly divided into two groups, small molecule drugs and biologics, and the small molecule drugs have dominated

the new drug approvals in each year from 2015 to 2018 [15, 16, 17, 18]. For example, 42 small molecule drugs and 17 biologics were approved in 2018, where the small molecule drugs accounted for 71% [15]. This section will focus on the challenges in developing new small molecule drugs.

The screening of promising chemical compounds, either through experimental or computational platforms, have been well developed and established in pharmaceutical industry for early drug discovery. The target-based discovery of small molecule drugs usually involves the following steps: target identification, compound screening, optimization and candidate selection [5, 6, 19]. Through improved understanding of a certain disease, key factors that affect disease pathology or progression are discovered and some of them have potential in interacting with small molecules to treat the associated disease. These key factors can be proteins, DNAs or RNA molecules and are referred to as targets [19]. At the step of compound screening, such as high-throughput screening, a chemical library is screened against the target to identify active chemical compounds, called hits. Lead compounds are identified by optimization and validation of the hits from screening methods to improve activity and selectivity [19]. The resulted leads are further improved on deficiencies such as off-target effects, and hopefully will lead to a drug candidate for preclinical and clinical research.

For screening of chemical compounds, there have been a variety of technologies to identify hits, e.g. high-throughput screening, structure-based design, fragment screening or virtual screening. Maraviroc,a chemokine receptor antagonist invented by Pfizer, is an instance of successful drug innovation through application of high-throughput screening [5]. The Pfizer chemical library, containing around 500,000 compounds, was screened against the CC-chemokine receptor 5 (CCR5). A weak agonist hit was identified by the screening tests and led to a series of structure-activity relationship studies. There were 4,000 and more compounds synthesized and optimized to achieve the desired antiviral activity [5, 20].

The major advantage of chemical screening methods is that they are faster than traditional methods by several orders in speed while maintaining reasonable costs [6]. It only takes several months to screen a set of one million chemical compounds with controllable quality [5]. However, the overall drug R&D has not displayed a turnover in efficiency since

the introduction of screening methods. The success rate of a drug candidate to pass all clinical trials remains stable, and the efficiency of drug R&D characterized by money spent on each approved drug also keeps declining.

The underlying assumption of screening methods is that good performance in screening tests, such as high binding affinity to certain targets, can lead to promising drug candidates that are effective in humans to treat the associated disease. However, it is risky to fully rely on this assumption for two reasons. One reason is that it requires much accurate understanding of the disease and identification of the potential targets [21, 5, 22, 23]. A good target with potential to be drugged against ought to be a key factor of disease pathology and not cause dramatic side effects. If a wrong target was identified, the hits obtained from screening tests will end up in failure eventually. The other reason is that screening tests naturally lack the context of the complex biological system. Assuming the targets are the right ones to be tested against, the performance of chemical compounds in screening studies, either good or bad, cannot be used as solid proof to determine whether a compound would be active and safe in human or it should be eliminated from further studies [6]. Ideally, all types of tests, including both screening and animial tests, should be conducted and the results be integrated to maximize the success rate of a hit leading to a drug candidate and the FDA approval. However, this is not economically feasible. There need more powerful approaches that can predict and integrate the information at systems level to aid drug discovery and development.

1.1.3 Lack of Understanding of Disease Mechanisms

As described in the section named "The potential risk in chemical screening methods for early drug discovery", screening methods are based on target identification, which requires deep understanding of disease mechanisms [21, 5, 22, 23]. And it has been explained that the quality of identified targets is crucial for success rate of screening methods. In addition, the section of "The cautious regulator and clinical trial failures" has explained that the primary cause of failures in Phase II or III is insufficient efficacy instead of drug safety. In the recent work by Cook et al, the research team analyzed the drug projects closed between 2005 to

2010 at AstraZeneca. They found that 40% of closed projects due to insufficient efficacy were claimed to lack data on a clear linkage of the target to the disease mechanism [21]. By contrast, projects with higher confidence in target validation or better understanding of the connection between the target and disease were less likely to fail due to efficacy issues [21].

However, understanding disease mechanisms and pathology remains challenging for some diseases, such as cancer and respiratory diseases [24, 25, 26, 27]. For example, drug development for influenza resulted illnesses are limited because of the lack of characterization of the immune response during influenza infection. For cancer, there have been major limitations of the animal models used in associated studies [25].

1.2 Challenges in Using High-throughput Biological Data to Reveal Disease Mechanisms

High-throughput biological data, including abundance data (e.g. mRNA level) and interaction data (e.g. PPI), has been growing exponentially in availability during the past two decades, fueling a rapid progression in systems biology research. Transcriptomes of single cells, tissues under different conditions, and disease states at multiple scales have been generated as a result of continually improved technologies [28]. In the past decade, thousands of papers about the two dominant measurements of transcriptional profiles (microarray and RNA-Seq) have been published per year [28].

Transcriptional profiling is a promising approach to characterize the regulatory behavior and dynamics of the immune response as well as the associated influence on physiopathology. In addition, time-series experiments of transcriptome analysis are able to characterize global dynamical change in a system and biological functions of genes which may indicate the relevant regulation pathways [29]. Different bioinformatics tools have been developed to analyze the transcriptomic data, [30, 31, 32, 33] many of which could also lead to clinical improvements. Potential applications of computational algorithms include faster, more accurate diagnosis of diseases, suggestion of personal and high-targeting treatment, and monitoring

of trends in immunopathology. Although these tools are designed for varying diseases, there may be overlap in strategy applicability.

By far most research has focused on identifying differentially expressed genes using transcriptomic data through statistical tests or inferring pathway activity by a variety of methods. However, there remain challenges in interpretation of complex tissue level or multiscale transcriptomic data. One significant confusion is the understanding of gene regulation. Expression signals of genes from disease-perturbed samples might increase or decrease during a period of time. But this might not be an indication that the genes have been activated or suppressed due to the perturbation, since changes of cellular composition in the biological sample might also induce the fluctuations of gene expressions. Possible composition change in cell population induced by diseases includes migration, proliferation, and differentiation. Transcript changes in cellular demographics and changes in gene regulation in temporal (e.g. time-point) samples should be treated separately. In addition, transcriptomic data has limited power when identifying disease associated genes or pathways. It has been shown that disease-related genes sometimes yield low gene expression levels [34] in the tissue or are less variated across time [35]. Therefore, more powerful and systems approaches are needed to better interpret the disease mechanisms.

1.3 Dissertation Overview

Advances in life science technologies have enabled researchers to explore exponentially growing biological datasets. These datasets include data ranging from the molecular level to cellular or tissue level. However, it remains a challenge to remove the noise and interpret the data to gain insights in disease pathology and drug mechanisms. In this work, we demonstrate how computational or statistical methods can assist in identifying disease or drug mechanisms.

Chapter 2 focuses on the statistical analyses of gene expression data of samples treated with four different types of influenza vaccine adjuvants through collaboration with RIKEN (Japan). Differentially expressed genes were identified by comparing the test samples with

the control, and the enriched functional annotations were analyzed to help understand the mechanisms of each adjuvant. The identified mechanisms provided meaningful insights and research direction in developing the novel adjuvant.

Chapter 3 takes various types of deconvolution algorithms to predict dynamic changes in immune cell quantities induced by influenza virus infection. There are two motivations for the prediction analysis. One is that changes in immune cell quantities, especially infiltration of certain immune cells to the lung, have been found to be relevant to deadly outcomes in disease pathology. The other motivation is that gene expression needs to be adjusted for changes in cell quantities to reveal differential expression caused by the perturbation of a disease. Identification of key disease-associated genes were much improved by using adjusted gene expression.

Chapter 4 shows that network topology predominantly determines the prediction accuracy of a drug target inference algorithm, ProTINA, published in former literature. Based on the findings, a novel target inference algorithm was proposed and showed better performance than ProTINA. This algorithm integrates protein betweenness and differential expression representing network topology and drug signatures, respectively, and is faster and more tractable than ProTINA. This chapter also discusses how the algorithm can be applied to different research scenarios. Chapter 5 focuses on the conclusions, accomplishments and potential future work of the previous chapters.

2.0 Comparing Gene Expression from Mice Treated with Different Influenza Vaccine Adjuvants

2.1 Introduction

Seasonal influenza infection is a threat for human beings. Induced by Influenza A (H1N1) viruses, the most recent global pandemic happened in 2009. Yearly influenza viruses could be lethal for people at high risk: people of 65 years old or older, people with chronic diseases, pregnant women or young children [36, 37]. According to Centers for Disease Control and Prevention, influenza A(H3N2) viruses and influenza B viruses circulated in the United States during the 2017-2018 influenza season, which caused an estimated 45 million influenza illnesses and 61,000 deaths [38]. Though the numbers vary, influenza viruses result in millions of sicknesses, hundreds of thousands of hospitalizations, and thousands or tens of thousands of deaths every year in the United States.

Infection with influenza raises challenges in clinical diagnosis and treatment. There are several antiviral drugs available which benefit people at high risk. But they usually work only if treatment starts within the first 2 days of illness. Most of these drugs are not available for young children, such as baloxavir marboxil (Xofluza). Vaccines remain the best strategy against the influenza virus infection.

Transcriptome analysis can help to characterize immune response or gene regulations induced by influenza vaccination and reveal immune mechanisms promoting vaccine efficacy and guide the design of new adjuvants. In this project, we obtained RNA-Seq gene expression data from mice treated with four different vaccine adjuvants from the collaborators at RIKEN (Japan). These adjuvants are AddaVax, aluminum precipitates (Alum), Complete Freund's Adjuvant (CFA) and CTA1-DD, where Alum and CFA serve as controls. Alum was invented 90 years ago and has been widely used as vaccine adjuvant for a variety of vaccines. Squalene-based AddaVax and cholera toxin-derived CTA1-DD have been developed in recent years. CFA is derived from inactivated and dried mycobacteria and currently is for research usage only. With our collaborators, we found that AddaVax is the most significant

and promising for future vaccine adjuvant development. Our functional annotation analysis found that much of the activity was associated with vitamin D_3 activity and provided future research direction in developing influenza vaccine adjuvant. This is also a demonstration of transcriptomic data and computational approaches in aiding vaccine development. Please note that this is a side project for training purposes, and in this chapter we focus on the statistical analysis of RNA-Seq gene expression data.

2.2 Methodology

2.2.1 Normalization and Statistical Analysis of RNA-Seq Data

The spleen and lymph node organs were extracted from the mice treated with vaccines containing one of four different types of adjuvants, namely AddaVax, Alum, CFA and CTA1-DD (where Alum and CFA serve as controls), and their mRNA levels were measured by RNA-Seq analysis. There were 3 biological replicates for each adjuvant or organ type.

The RNA-Seq reads were obtained from Dr. Hidehiro Fukuyama at RIKEN (Yokohama, Japan), and the normalization or statistical analysis were performed by the CLC Genomics Workbench (version 10.0.1). The reads were mapped to the annotated mouse genome, "Mus musculus ensembl v80 Sequence", and transcripts per million (TPM) were calculated as explained by Li et al. [39]. TPM counts were normalized by the trimmed mean of M values (TMM) method [40] to calculate effective library sizes. The count per million (CPM) values were calculated for each gene by using the effective library sizes obtained from TMM method, and log2-scaled CPM values were utilized for principle component analysis (PCA). For differential expression analysis, each gene was modeled by a Generalized Linear Model (GLM) assuming negative binomial distribution, and fold changes were calculated accordingly. The p-values were obtained by performing Wald test.

2.2.2 Functional Annotation Enrichment Analysis

Differentially expressed genes (fold change > 1.5, and p-value < 0.05) were obtained as explained in the section above. Their significant functional annotations were obtained from DAVID [41, 42] (enrichment score > 2).

2.3 Results

2.3.1 Identifying Differentially Expressed Genes in RNA-Seq Data from Influenza-infected Samples Pre-treated with Various Adjuvants

Through differential expression analysis, we found lymph node samples are more significant than spleen samples, and AddaVax is the most significant and promising adjuvant for future development (Figure 2). Using Alum as control, 475 and 630 genes are differentially expressed in samples treated with CTA1-DD and AddaVax, respectively (fold change > 1.5, and p-value < 0.05). If CFA is used as control, there are 13 and 430 differentially expressed genes for CTA1-DD and AddaVax, respectively, under the same thresholds for fold change values and p-values. Therefore, AddaVax is the most significantly different from control adjuvants (Figure 3).

As the comparisons by using Alum as control have shown more significant results (Figure 2), we focused on differentially expressed genes from CTA1-DD or AddaVax versus Alum for functional annotations. The enrichment scores of significant genes' functional annotation clusters were obtained from DAVID [41, 42]. These genes are mostly enriched in immune response (biological process), binding (molecular function) and extracellular region (cellular component) as shown in Table 1. Especially, significant genes of AddaVax are enriched in collagen, lectin and steroid biosynthesis 1. This has provided inspiration for our collaborators to exploit associated biological processes and find that the mechanism of AddaVax is associated with vitamin D_3 activity. A vitamin D_3 analog adjuvant, oxarol, has been reported in the recent publication [3], and our collaborators are preparing for another article on associated research.

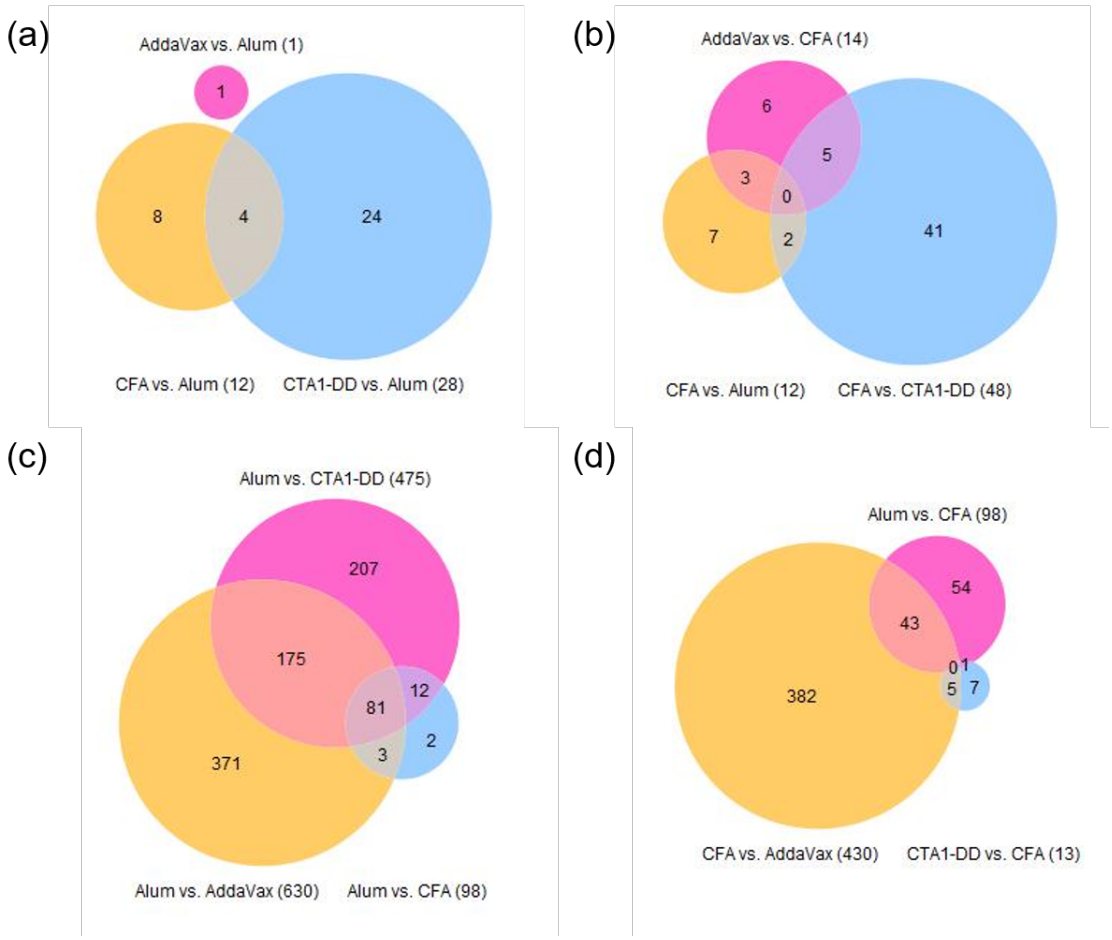


Figure 2: The number of differentially expressed genes in spleens (a)(b) or lymph nodes (c)(d). In (a)(c), Alum serves as the control, and CFA is the control in (b)(d).

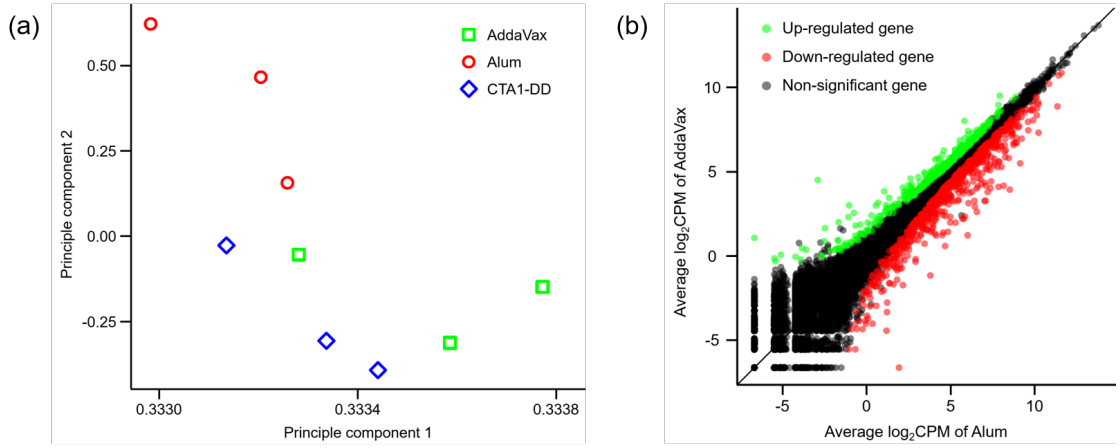


Figure 3: (a) CPMs of 3 adjuvant samples (AddaVax, Alum and CTA1-DD) are log2-scaled and averaged respectively. The first two principle components are shown here, color coded by adjuvant type. (b) Comparison between average log2-scaled CPMs of AddaVax and Alum. Up-regulated significant genes (FDR < 0.25) and down-regulated significant genes are colored green and red.

Table 1: Significant genes (FDR < 0.25) in comparison between AddaVax and Alum samples are analyzed by DAVID for annotations. Annotation terms are summarized with referring enrichment scores listed in the table.

Annotation Cluster	Enrichment Score
Immune response to virus	33.9
Signal peptide	25.9
Cellular response to interferon	11.4
Collagen	6.6
Lectin	6
Chemotaxis	5.2
Double-stranded RNA binding	5.1
Endoplasmic reticulum	4.7
Immunoglobulin domain	4.6
Zinc finger region	4.1
Double-stranded DNA binding	4.1
Steroid biosynthesis	3.4

3.0 Predicting Host Immune Cell Dynamics and Key Disease-Associated Genes Using Tissue Transcriptional Profiles

3.1 Introduction

Accurately identifying and quantifying the immune cells is critical to understanding both how the body manages disease and how immune mismanagement may increase the overall disease pathology (e.g., immunopathology). The behavior of immune cells is a primary factor in the overall disease pathology [43, 44, 45, 46, 47, 48, 49]. The immune response is a complex process that coordinates the activation, migration and differentiation of a large variety of immune cells [50, 51, 52, 53]. A very common factor of disease pathology is overly aggressive or dysregulated immune responses in which diseased tissues are observed to have abnormally high numbers of inflammatory immune cells. A specific example is lethal influenza infections, which are often characterized by extremely high quantities of macrophages or neutrophils that infiltrate into lungs [54, 55, 56, 57]. It has also been shown in influenza infection studies that modulating inflammatory immune cell counts by interfering with immune cell trafficking or activation can significantly improve infection outcomes [58, 59, 60]. An accurate quantification of immune cells is essential to identifying the mechanisms of disease pathology and can provide insights in innovating treatments.

While cell count data is necessary to mathematically model disease development, such data is often limited and not nearly as accessible to the research community as genomics data. Fluorescence-activated cell sorting (FACS) is one of the most common methods to quantify cells in a sample. However, FACS requires significant amount of tissue for analysis, which complicates the design of experiment [61]. Moreover, FACS data repositories are not yet well established, although ongoing efforts such as ImmPort [62] are aiming to improve this. Gene expression data on the other hand is widely available in curated repositories, such as GEO [63]. To support identifying the mechanisms behind disease pathology and promote mathematically modeling the complex systems linking disease and immune responses, it

would be a major benefit to be able to exploit gene expression data to identify and count the immune cells in a sample.

The computational challenge is to use the changes in the number of RNA transcripts within a tissue that is caused by the changing numbers of immune cells to infer, i.e., count, the number of immune cells themselves. Both the signaling pathways activated by a disease and the increased localization of immune cells result in changes in the number of RNA transcripts within a tissue (see Figure 4) [64, 29]. Most genomics research focuses on identifying differentially expressed genes by comparing gene expression from diseased tissue samples with the control. However, given how immune cell infiltration impacts RNA transcript counts, one should be able to infer the changes in numbers of immune cells by examining the expression data. It is also significant to consider that without adjusting for changes in cellular composition, the obtained differentially expressed genes may include false positives that are not related to regulation activities but due to changes in cell populations. Computational approaches based on gene expression data, i.e., deconvolution algorithms, have been developed to address these issues and can assist in the identification of gene regulations.

Deconvolution algorithms attempt to quantify cell counts in a mixed sample by using gene expression data [65, 66]. The expression profiles that can be analyzed by deconvolution algorithms include tissue samples collected from animals (e.g., mouse lung tissue samples infected by influenza viruses [29]) and clinical samples from patients (e.g., blood samples from systemic Lupus erythematosus patients [67]). Figure 4b provides a summary of the algorithms used in this study. Two general strategies have been proposed. One takes a bioinformatics approach and uses longitudinal clustering of time-series data to isolate sets of genes and then associates the gene sets with candidate cell types using statistical tests, such as cell type enrichment (CTen) [64, 29]. In this approach, the gene expression patterns are assumed to be correlated with dynamic changes of enriched cell types. This strategy has advantages: firstly, no prior knowledge of the cellular composition in the test sample is required; secondly, all time points are considered simultaneously, and the method can easily be extended over many experimental conditions.

The other common strategy postulates that tissue can be modeled as a linear combination of gene expression profiles derived from pure cell populations. In general, these approaches

assume

$$\boldsymbol{A} \cdot \boldsymbol{x} = \boldsymbol{b}, \tag{3.1}$$

where $\boldsymbol{A}$ is a $n \times p$ matrix of expression intensities of p kinds of pure cells for n genes [65, 68], $\boldsymbol{b}$ is a vector of gene intensities in the test sample, and $\boldsymbol{x}$ is a vector of proportions for cell types in the matrix $\boldsymbol{A}$. Since n is usually larger than p, the linear problem is overdetermined and different regression strategies are applied in various algorithms. Cell fractions are solved through simulated annealing [65, 68], bounded linear least-squares regression (LLSR) [67], quadratic programming in digital sorting algorithm (DSA) [66], elastic net regularization in digital cell quantification (DCQ) [69], and a modified support vector regression (v-SVR) in a tool named the cell-type identification by estimating relative subsets of RNA transcripts (CIBERSORT) [70]. These algorithms require information of cell composition in the sample and reference gene expressions of pure cells. Both requirements can limit their application in inference.

An unexamined question is how accurate these deconvolution algorithms are for predicting cell quantity changes in animal-derived tissue sampled over time (i.e., dynamic data) and how to utilize their predictions to improve the identification of gene regulation related to the disease. Most deconvolution algorithms are only tested by simulated expression data [71], or data from artificial samples comprised of three or four distinct cell/tissue types [66, 67]. Seldom is algorithm performance examined by in vivo collected tissue samples, such as whole blood samples [67]. To better understand performance, we evaluate four representative deconvolution algorithms: CTen [64], modified LLSR (referred to as MLLSR in the following) by Abbas et al. [67], CIBERSORT [70] and DCQ [69] (Figure 4b), by a common microarray dataset of influenza-infected lung tissue sampled at multiple time points [29]. Estimates from each deconvolution algorithm are compared with cell count data measured by FACS under the same experimental conditions. We then propose a new approach to identifying significant genes associated with disease by application of predictions from deconvolution. We find that the ranking is significantly increased for important genes known to be factors of virus replication and disease pathogenesis, after adjusting the bias of differential expression analysis due to immune cell infiltration. Lastly, we conclude with a discussion on how cell

count inference algorithms can be incorporated into analytical pipelines to improve disease factor discovery.

3.2 Methodology

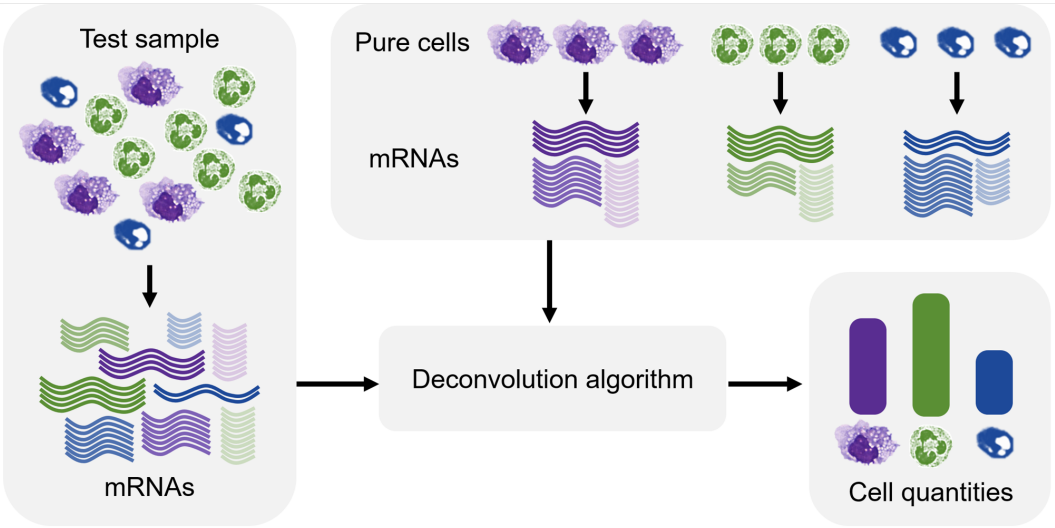
3.2.1 Ethics Statement

All mouse experiments were performed following the University of Tokyo's Regulations for Animal Care and Use, which are approved by the Animal Experiment Committee of the Institute of Medical Science, University of Tokyo (approval number: PA10-13). All experiments involving H5N1 virus were performed in biosafety level 3 containment laboratories at the University of Tokyo, with the approval by the Ministry of Agriculture, Forestry, and Fisheries, Japan.

3.2.2 Microarray Analysis of Mouse Lung Tissue

Complete and detailed methods for infection, tissue collection and tissue treatment to perform lung gene expression analysis for the mice used in this study are reported in Shoemaker et al. [29]. Briefly, 42 mice per cohort were inoculated with 10^5 plaque-forming unit per gram of lung (PFU) of the A/Kawasaki/UTK-4/09 H1N1 virus (H1N1), A/California/04/09 H1N1 virus (pH1N1), or the A/Vietnam/1203/04 H5N1 virus (H5N1). A cohort of animals mock-infected with PBS (phosphate-buffered saline) served as the control for a total of 168 mice. At 14 time points, three mice per cohort were humanely sacrificed, their lungs harvested, and the left-lower section used for gene expression analysis by single-color microarray (the remaining sections were used for cytokine assay and Western blot analysis). Data was background corrected and then quantile normalized using the "limma" R package [72] and are available in the gene expression omnibus (GEO) repository (GSE63786).

(a)



(b)

Deconvolution Algorithm	Information Of Pure Cell Gene Expression	Type Of Data Required	Mathematical Operations	Outputs
CTen	HECS (a built-in gene marker database)	Time course, log fold change gene expression	Data is clustered (e.g. WGCNA [31]) and Fisher's exact test is applied	Enrichment scores and weighted log fold change of cell counts
MLLSR	User-provided gene expression of pure cells	Gene expression intensity	Linear least-squares regression	Cell fractions
CIBERSORT	User-provided gene expression of pure cells	Gene expression intensity	Modified support vector regression (ν-SVR)	Cell fractions
DCQ	A compendium of immune cell surface markers and their gene expression (a built-in database)	Normalized log fold change gene expression	Elastic net regulation	Relative cell quantities

Figure 4: Overview of deconvolution algorithms. (**a**) Gene transcript counts change as the cellular makeup of a sample changes. Deconvolution algorithms postulate that the change in the cellular makeup of the tissue can be inferred from the tissue's gene expression by exploiting the transcriptional profiles of pure cells. (**b**) Inputs, mathematical operations and outputs of the four deconvolution algorithms reviewed in this paper.

3.2.3 Flow Cytometry

Five mice per time point per cohort were infected with 10^5 PFU of virus. Five uninfected (naïve) mice served as the negative control. Lungs were collected from mice and incubated with Collagenase D (Roche Diagnostics; final concentration: 2 g/mL) and DNase I (Worthington; final concentration: 40 U/mL) for 30 min at 37 °C. Single-cell suspensions were obtained from lungs by grinding tissues through a nylon filter (BD Biosciences, San Jose, CA, US). Red blood cells (RBCs) in a sample were analyzed using an RBC lysing buffer (Sigma-Aldrich, St. Louis, MO, US). Samples were resuspended with PBS containing 2 mM EDTA and 0.5% bovine serum albumin (BSA), and the cell number was determined using a disposable cell counter (OneCell, Fine Plus International, Kyoto, Japan). To block nonspecific binding of antibodies mediated by Fc receptor, cells were incubated with purified anti-mouse CD16/32 (Fc Block, BD Biosciences). Cells were stained with appropriate combinations of fluorescent antibodies to analyze the population of each immune cell subset (see Table 2). The following antibodies were used: anti-CD49b (DX5: BD Biosciences), anti-FcεRI (MAR-1: eBioscience), anti-c-kit (2B8: BD Biosciences), anti-CD45 (30-F11: eBioscience), anti-CD11b (M1/70: BD), anti-CCR3 (83101: R&D), anti-F4/80 (BM8; eBioscience), anti-CD11c (HL3: BD), anti-Gr-1 (RB6-8C5: BioLegend), anti-NK1.1 (PK136: BD Biosciences), anti-B220 (RA3-6B2: BD Biosciences), anti-CD3ε (145-2C11: BD Biosciences), anti-CD4 (RM4-5: BioLegend), anti-CD8α (53-6.7: BioLegend), and CD69 (H1.2F3: BD Biosciences). All samples were also incubated with 7-aminoactinomycin D (via-probe, BD Biosciences) for dead cell exclusion. The data of labeled cells were acquired on a FACSAria II (BD Biosciences) and analyzed with FlowJo software version 9.3.1 (Tree Star). To evaluate the statistical significance of cell counts across time, we computed false discovery rate (FDR) values by comparing cell counts at each time point and uninfected (naive) mice.

3.2.4 Prediction of Cell Fractions by MLLSR, CIBERSORT and DCQ

Microarray data of 17 cell types were gathered to construct a library of gene expression profiles from populations of a single cell type for deconvolution by MLLSR and CIBERSORT. These cell types included: B cell, Kdo (12h) stimulated B cell [73], naïve CD4+ T cell,

natural CD4+ regulatory T cell [74], resting naïve CD8+ T cell, resting memory CD8+ T cell, stimulated naïve CD8+ T cell, stimulated memory CD8+ T cell [75], immature dendritic cell (imDC), mature DC (maDC), a unique subset of regulatory DC (sDC), IL-10 treated sDC [76], lung [29], macrophage [77, 78], LPS (6h) stimulated macrophage [78], monocyte [79], and NK cell [80]. Data was background corrected (monocyte data not corrected due to unavailability) and then quantile normalized. Lowly-expressed genes were excluded (intensity > 256 or log-scaled intensity > 8) based on the assumption that large intensity values benefit identification of various cell types and deconvolution of gene expression reference. Gene markers of a certain cell type in the library were selected by sorting the ratio of the intensity of a gene for the cell type divided by the average of the same gene for all other cell types. Top 100 genes with the highest ratios of each cell type and their expression intensities were collected as information of pure cells that serve as matrix $\boldsymbol{A}$ in regression. The intensities of these gene markers in an infected lung tissue sample served as the input as $\boldsymbol{b}$ to MLLSR and CIBERSORT. MLLSR was rebuilt in R according to the literature. Equation (1) is solved by the R function 'lsfit' and the minimum negative coefficient is removed until there is no negative value [67]. Analysis by CIBERSORT was implemented online with settings as default. It uses ν-SVR, which minimizes both a loss function and a penalty function. CIBERSORT contains a feature selection step, in which less-variated genes are discarded to reduce overfitting. In the last step, negative coefficients were set as zeros and all coefficients are scaled to sum of one [70].

DCQ does not require or accept user-defined information of pure cells, and it depends on an immune cell compendium that consists of a collection of 61 cell surface markers for 223 diverse cell types (213 of them are immune cells) and expression profiles of these cell subsets obtained from ImmGen Project [81]. Intensities of all genes available in lung tissue microarray data were uploaded to DCQ. As suggested [69], preprocessing of the data was performed by log-scaling and subtracting the control sample, and every entry was divided by the global standard deviation. The parameters, number of repeats and lambda.min, were left as default.

3.2.5 Gene Co-Expression Analysis by WGCNA and Cell Enrichment Analysis by Cten

Microarray datasets of mouse lung tissues at 14 timepoints (0, 3, 6, 9, 12, 18, 24, 30, 36, 48, 60, 72, 120, and 168h) after infection by H1N1, pH1N1, and H5N1 were obtained from literature [29] (methods summarized above). For each strain, a gene was differentially expressed if it was significant for at least one time-matched comparison with mock samples (fold change > 2; FDR < 0.01). Log fold change values of these differentially expressed genes were utilized for the construction of the co-expression network by the WGCNA package [82] in R. Block-wise network construction was implemented for pH1N1 and H5N1 due to the large size of their datasets. A soft-thresholding power of 8, 9, and 8 was set respectively for H1N1, pH1N1, and H5N1, based on scale-free topology fitting (Figure 17). Modules were not merged (height cut = 0) for all strains, because merging modules can generate lower correlations among genes within a module [29]. Module eigengenes and module membership of genes were calculated as well. For each module, genes positively or negatively correlated to the eigengene were separated into two submodules, according to signs of their memberships. Negative submodules are denoted by an extra minus sign.

Submodules of genes obtained from WGCNA were uploaded to CTen for detection of cell enrichments [64]. Genes from each module were compared with CTen's gene marker database. Enrichment scores were computed as $-\log_{10}(p\text{-value})$ using p-value from Fisher's exact test. When more than one submodule was annotated by the same cell subset, the submodule with the highest enrichment score was chosen to represent this cell type.

3.2.6 Comparison between Estimated Cell Quantities and Flow Cytometry Data

To explore the accuracy of each algorithm, we calculated R^2 and normalized mean squared error (NMSE) by comparing normalized predictions with processed cell counts from flow cytometry. NMSE is defined as the following equation:

$$\text{NMSE}(\boldsymbol{x}, \boldsymbol{y}) = \frac{\|\boldsymbol{x} - \boldsymbol{y}\|^2}{\|\boldsymbol{x}\|^2}, \tag{3.2}$$

where $\boldsymbol{x}$ is a vector of normalized predictions, and $\boldsymbol{y}$ is a vector of processed cell counts. Although animals used for flow cytometry are not the same as those for microarray analysis, mean values of both predicted and measured cell quantities are used to account for errors due to diversity of animal samples for all comparisons.

For MLLSR and CIBERSORT, log fold change of estimated cell fractions at each time point versus hour 0 was calculated and compared with log fold change of cell counts to compute R^2 and NMSE scores. The total number of all cells in the lung tissue is not available because large cells (e.g., epithelial cells) were filtered before cell sorting. We assumed the total number of cells did not significantly change in the seven days following infection (proof shown in Figure 18). Therefore, the log fold change of estimated cell fractions (a test sample vs mock sample) should approximately equal to the log fold change of cell counts. Because the physical meaning of the output from DCQ is unclear, we compared it with three kinds of measurement: (i) log fold change of cell counts, (ii) change of cell counts, and (iii) change of normalized cell counts (the cell number of a cell type divided by the number of total live cells). The predicted relative cell quantities from DCQ for more than 200 cell subsets were summed up according to the annotations of its immune cell compendium. For CTen, the expression pattern of a gene marker is assumed correlated to the dynamic change of the referring cell type, and eigengene profiles were normalized then compared to cell count data. The eigengene profiles of the chosen modules were scaled as 0–1, and the profiles from negative submodules were multiplied by -1 before scaling. The R^2 and NMSE scores were computed in comparison between the normalized eigengene profiles and log fold change of cell counts.

3.2.7 Computation of Adjusted Gene Expression and Identification of Significant Genes

Based on the linear relation i.e., Equation (1) applied in deconvolution algorithms, CIBERSORT-adjusted gene expression $\boldsymbol{g}$ was defined as the following:

$$\boldsymbol{g} = \boldsymbol{b} - \boldsymbol{A} \cdot \hat{\boldsymbol{x}}, \tag{3.3}$$

where $\boldsymbol{A}$ is a $n \times p$ matrix of expression intensities of p kinds of pure cells for n genes, $\boldsymbol{b}$ is a vector of expression intensities in a given sample, and $\hat{\boldsymbol{x}}$ is a vector of estimated cell fractions for this sample using CIBERSORT. Similarly, cell count-adjusted gene expression $\tilde{\boldsymbol{x}}$ was defined as

$$\boldsymbol{e} = \boldsymbol{b} - \boldsymbol{A} \cdot \tilde{\boldsymbol{x}}, \tag{3.4}$$

where $\boldsymbol{b}$ and $\boldsymbol{A}$ are the same with the above while $\tilde{\boldsymbol{x}}$ is equal to immune cell counts measured by FACS divided by the average total number of cells per mouse lung [83]. For each gene from influenza-infected microarray data, the adjusted gene expression $\boldsymbol{g}$ and $\boldsymbol{e}$ was quantile normalized and compared with that of mock data per time point per sample cohort by the R package limma [72], and the FDR was computed. Standard differential expression (DE) analysis was performed by comparing gene expressions of infected samples to mock samples per time point per cohort and FDR values were computed. For both methods (adjusted expression and standard DE), the minimum FDR values among all time points was utilized to characterize the significance of the associated gene. Genes with FDR values less than 1×10^{-4} for at least two time points were analyzed by DAVID [41, 42] for functional annotations.

3.3 Results

3.3.1 Dynamic Change of Immune Cell Quantities Induced by Influenza Infection

To characterize the accuracy of predictions from the set of deconvolution algorithms, the number of immune cells in mouse lung at five timepoints (day 0, day 1, day 2, day 3, and day 7) after influenza virus infection were measured by FACS. The cohorts of H5N1 and pH1N1-infected animals had higher immune cell counts (Figure 19), which was consistent with studies of lungs infected with highly pathogenic viruses [54, 55, 56]. Figure 5 shows the number of select immune cells within the lung overtime (Figure 19 shows the results for all other immune cell counts measured). Macrophages were significantly infiltrated into

the mouse lung after day 2 for all sample cohorts (Figure 5). B cell counts did not increase significantly from the counts observed in mock animals until day 7 for the sample cohorts of H1N1 and pH1N1, while for H5N1 the quantity of B cells shows a significant decrease at both day 2 and day 7. For T cells, the H1N1 cohort exhibited an increase in cell number on day 7, whereas a large increase occurred much earlier (since day 2) for pH1N1. However, for H5N1 the variation of T cell counts was insignificant. We observe that CD4+ T cells show similar trends across time: the cell counts did not significantly increase until day 7 for the H1N1 and pH1N1 cohorts, while the cell counts for H5N1 are relatively stable (Figure 19). Dendritic cells (DCs) for the cohort of H1N1 are up-regulated at day 7, and for pH1N1 as well as H5N1 the cell counts greatly increased since day 2 (Figure 19). NK cell quantities kept increasing starting from day 3 for H1N1, and day 2 for pH1N1 and H5N1. These dynamic profiles of cell counts obtained from FACS were used to evaluate cell count predictions from the four algorithms.

3.3.2 Cibersort More Accurately Predicts Quantity Changes of T Cells and Macrophages than MLLSR

As explained above, MLLSR and CIBERSORT utilize expression intensities of pure cells and cell mixture samples to calculate cell fractions by linear regression. To improve computational efficiency, they both recommend using expression intensities of user-selected cell marker genes instead of the whole genome. The major difference between the two algorithms is the regression tool they use (see Materials and Methods). Here we evaluate the accuracy of MLLSR and CIBERSORT in predicting the mean cell counts observed in H1N1, pH1N1 or H5N1-infected lung. Two performance measurements for accuracy were provided: the R^2 values and NMSE. As paired data from the same animal is not available (see Discussion on data limitations), we evaluate each algorithm's ability to accurately predict the mean population observed (Materials and Methods).

MLLSR failed to capture average changes in immune cell populations while CIBERSORT demonstrates significantly better accuracy (Figure 6). Predicted fractions from MLLSR have comparatively low R^2 values and large NMSE scores for most cell types while CIBERSORT

has improved R^2 and reduced NMSE. MLLSR's predictions overestimated the fractions of macrophages to be more than 100% at several time points (Figure 6a). Estimations of macrophages from CIBERSORT had high R^2 values: 0.59, 0.95 and 0.95 for H1N1, pH1N1 and H5N1 respectively, with acceptable NMSE scores (0.59, 0.17, and 0.26), but similar to MLLSR, we observed that CIBERSORT had a tendency to predict higher macrophage counts than those measured (Figure 6b). CIBERSORT's estimations on T cells, activated CD8+ T cells and DCs (Figures 6b and Figure 20) fit well with cell count data for select sample cohorts. For example, R^2 values of T cells was 0.86 and 0.97 for H1N1 and pH1N1 respectively while that of H5N1 was 0.25 (Figure 6b). When predicting less abundant cell types, MLLSR is unable to estimate CD4+ T cells and CIBERSORT fails to estimate CD8+ T cells (Figure 21). The R^2 values across all cell types and all cohorts for MLLSR was 0.13 and 0.34 for CIBERSORT (Figure 22).

Next, we analyzed the time course trajectories of predicted cell fractions. Although corresponding cell count data was lacking for the majority of time points which have gene expression data, the smoothness of the time-course cell quantity curve and the timescales associated with changes in cell fractions can provide another measurement of inference quality. Figure 7 displays estimated variation of cell populations across time. We observed that different from cell counts measured by FACS, predictions from MLLSR and CIBERSORT for B cells variate dramatically between 0 and 0.03 for at least one sample cohort (Figures 7a,b and 5). Both algorithms failed to capture the decrease of B cells for H5N1, and predictions from MLLSR are mostly unchanged throughout the seven-day time frame. The increase of estimated macrophage fractions predicted using MLLSR and CIBERSORT is minimized for the H1N1 cohort than that of pH1N1 and H5N1, which is consistent with cell count data (Figures 7a,b and 5). However, the estimations for pH1N1 by MLLSR reached a peak at day 2 and then decreased, and the estimated cell fractions of macrophage for H5N1 showed a quick increase and stayed at a high level beginning at day 1. All of these disagreed with cell count data (Figure 5). CIBERSORT's performance in predicting macrophage cell counts is slightly improved relative to MLLSR as it better estimates the increase of macrophages for H5N1. For T cells, MLLSR and CIBERSORT accurately predict the increase of T cells for the H1N1 and pH1N1 cohorts, including that pH1N1 cohort has the highest increase among

all cohorts, while they both fail to predict the steady behavior of H5N1 (Figures 7a,b and 5). For other immune cell types, MLLSR estimates DCs to be almost unvaried and close to zero across the time points sampled for H1N1 (Figure 23). It also estimated CD4+ T cells and NK cells to be zero for almost all time points of all cohorts (Figure 23). Although this was corrected in predictions by CIBERSORT (Figure 24), both MLLSR and CIBERSORT cannot capture the rapid increase of DCs for the pH1N1 cohort. In general, CIBERSORT was more sensitive to the dynamic changes of immune cell quantities using gene expression profiles than that of MLLSR, and predicted time-course cell fractions from CIBERSORT were smoother than those of MLLSR.

3.3.3 DCQ Correctly Predicts Relative Cell Quantities of B Cells and Macrophages for the pH1N1 and H5N1 Cohorts

DCQ assumes that there is a linear relation between gene expressions and cell quantities in the same way as MLLSR and CIBERSORT, though it differs from them in three ways: (i) the input data were prepared by comparing a test sample to a reference sample, and the output was the relative change in cell quantities instead of the actual cell proportions in the sample; (ii) DCQ depends on its built-in immune cell compendium for regression and thus lacks flexibility; (iii) relative cell quantities were predicted using elastic net regulation (See Materials and Methods and Figure 4b). Since DCQ predicts the relative change of cell quantity, we compare the estimated relative cell quantities with FACS-measured cell count data that are preprocessed in three different ways (Materials and Methods). We found that adjusting the cell counts by the total live number of cells provided the best fits when using DCQ (Figure 22). The following results (including R^2 and NMSE values) were based on this comparison.

Capable of predicting a variety of immune cell populations, DCQ performs well on subsets of the sample cohorts. As shown in Figure 8, its predictions of B cells show high R^2 values of 0.84 and 0.92 for the pH1N1 and H5N1 cohorts (NMSE = 0.19 and 0.12, respectively). However, the R^2 for H1N1 was 0.16 (NMSE = 1.34). For macrophages, the estimated relative cell quantities agree well with normalized cell counts (especially for pH1N1, $R^2 = 0.98$, NMSE

= 0.11), while the estimations of H1N1 have lower accuracy (R^2 = 0.46, NMSE = 0.63). Analogous to MLLSR and CIBERSORT, DCQ also excessively estimates the quantities of macrophages for the H5N1 cohort (Figure 25). Regarding less abundant cell types such as natural killer T cells (NKT cells), DCQ was unable to correctly quantify its change of cell quantities using lung tissue microarray data. The overall R^2 value calculated by comparing DCQ's estimations to normalized cell count data was 0.32 for all sample cohorts.

Then we assessed DCQ's estimations in the time course using the trend in cell quantity variation. DCQ successfully captures the observed decrease in B cells around day 2 in H5N1 but fails to detect the increase at day 7 for the H1N1 and pH1N1 cohorts (Figures 7c and 5). For macrophages, DCQ outperforms CIBERSORT and accurately predicts the continuous increase starting from day 2 for the pH1N1 and H5N1 samples (with the exception of day 7 of H5N1). The stable behavior observed in T cell quantities is correctly predicted for the H5N1 cohort, while the rising accumulation of T cells after day 2 in pH1N1 samples is not captured. Additionally, DCQ failed to predict the apparent increase of neutrophils beginning at day 2 for H5N1 and the behavior of DCs for any sample cohort (Figures 19 and 26). In conclusion, DCQ is an effective tool for predicting select immune cell subsets and does not guarantee good performance for all samples.

3.3.4 CTen Shows High Accuracy When Predicting Dynamic Changes in Macrophages and Neutrophils for All Sample Cohorts

CTen, different from the aforementioned algorithms, did not predict cell quantities by linear regression. Instead, the approach suggests clustering the gene expression data and using overlap/enrichment to identify significant associations between clusters of genes and cell types. Here, we employed WGCNA to cluster genes with highly correlated expression [82]. The eigengene (the first principle component) of each cluster was then used to quantify the mean change in the cell counts. The normalized eigengene profiles were not expected to have the same magnitude as cell count data, but NMSE scores are still provided as a reference with R^2 values.

CTen shows consistent performance among samples for most cell types tested. It performed well on all sample cohorts for macrophages and neutrophils. Normalized eigengene profiles of macrophages for each cohort show high R^2 values compared to measured cell counts (Figure 8): that of the H1N1 cohort is 0.84 (NMSE = 0.59), pH1N1 was 0.88 (NMSE = 8.89), and H5N1 was 0.80 (NMSE = 5.96). CTen's estimations for neutrophils outperform DCQ with the H1N1 cohort $R^2 = 0.70$ (NMSE = 0.40), pH1N1 cohort $R^2 = 0.96$ (NMSE = 3.99), and H5N1 cohort $R^2 = 0.86$ (NMSE = 3.65). The R^2 values for H1N1, pH1N1 and H5N1 in NK cells were 0.59, 0.85, and 0.51, respectively, and in DCs were 0.41, 0.86, and 0.56, respectively (Figure 27). While the estimations for NK cells and DCs have relatively lower R^2 values, no cohort returns an extremely poor R^2 value. However, the normalized eigengene profiles of B cells did not fit well with cell count data (Figure 8), with an R^2 value of 0.03 for H1N1 (NMSE = 1.18), 0.52 for pH1N1 (NMSE = 0.68) and 0.40 for H5N1 (NMSE = 5.78). The R^2 value of CTen's predictions across all samples was around 0.37, which was the highest of all four algorithms.

The trend in cell quantity change predicted by CTen aligned with FACS data for select cell types. CTen successfully captured the increase of macrophages at days 2 and 7 for the H1N1 cohort, as well as the increase at days 2 and 3 followed by the decrease at day 7 for H5N1 (Figure 7d). Within the pH1N1 cohort, the macrophage quantities were predicted to reach a peak before day 3 followed by a slight decrease, a trend which disagrees with cell count data. For neutrophils, CTen correctly estimated the increase at days 2 and 3 followed by a slight decrease at day 7 for pH1N1, and the constant increase for H5N1 though the estimated decrease at day 7 was inaccurate (Figure 28). For the H1N1 cohort, the increase of neutrophil quantities was overestimated compared to FACS data. Additionally, CTen vastly improved predictions of dynamic changes in NK cells and DCs. Predicted dynamic changes in CD4+ T cells were the same with those in CD8+ T cells as shown in Figure 7d. Based on measured cell counts, the quantities of both cell types did vary at a consistent pace (Figure 5), while the variation was not consistent with what CTen predicts. For example, CD4+ T cells and CD8+ T cells are relatively constant across time for the H5N1 cohort where CTen's predictions show more deviation.

3.3.5 Improved Disease-Associated Gene Identification by Adjusting for Cellular Composition

One of the most important applications of deconvolution was to remove gene transcript count changes due to cell count changes, which should improve identification of gene expression activity associated more specifically with the disease being studied. Transcriptional profiles of diseased tissue varied over time as a result of both the fluctuation of immune cell populations (such as infiltration of macrophages and neutrophils) and the changes of the gene regulatory networks (activation or repression of certain genes). Based on this assumption, we demonstrated that deconvolution can be combined with statistical analysis to improve identification of disease-associated genes (see Materials and Methods). Briefly, changes in transcript levels due to changes in the cellular composition of the tissue were subtracted from the gene expression data. The resulting gene expression, represented by $\boldsymbol{g}$ (adjusted by predictions from CIBERSORT) or $\boldsymbol{e}$ (adjusted by cell counts from FACS), was then analyzed with established microarray statistical tools (e.g., "*limma*" [72]) to determine significant genes. The FDR values and associated ranks of genes generated using this adjusted gene expression were compared to those obtained through standard differential expression (DE) analysis of the unadjusted data.

Adjusted gene expression, based on either inference by CIBERSORT (i.e., $\boldsymbol{g}$) or measured cell counts (i.e., $\boldsymbol{e}$), largely improved FDR values and ranking for genes involved in influenza infection and the anti-viral immune response. Among the top 10 ranked genes obtained by sorting FDR values calculated from adjusted gene expression $\boldsymbol{g}$, *Psme1* was the most significant gene for all three sample cohorts (as shown in Figure 9). *Srp14* is ranked 7th and 9th for the H1N1 and pH1N1 cohorts, respectively (ranked 16532th for H5N1). The human PSME1 protein interacts with influenza A virus protein neuraminidase (NA) according to the protein-protein interaction (PPI) database VirHostNet2 [84]. The human SRP14 protein interacts with influenza nucleoprotein (NP), polymerase acidic protein (PA), non-structural protein 1 (NS1), polymerase basic protein 2 (PB2), RNA-directed RNA polymerase catalytic subunit (PB1), hemagglutinin (HA), NA, and the matrix protein 1 (M1) [84]. *Sqstm1*, *Mapkapk2*, and *Lamtor2* are also found in the 20 most significant genes for different cohorts

(Figure 9), all of which are well known for being associated with the NF-κB signaling [85] and MAPK/ERK pathways [86, 87]. However, in the standard DE analysis based on original gene expression, *Psme1* is ranked 456th, 1353th and 1145th for H1N1, pH1N1 and H5N1, respectively. *Srp14* is not included in the top 1000 genes with much higher FDR values. For *Sqstm1*, *Mapkapk2*, and *Lamtor2*, genes which are considered important in adjusted gene expression analysis, their rank is significantly lower than those from adjusted gene expression (Figure 9). Instead, *Cxcl9*, the chemokine that attracts NK cells and T cells [88], is ranked higher using unadjusted gene expression data (2nd for the pH1N1 cohort). For significant genes obtained from adjusted gene expression $\boldsymbol{e}$, which is computed based on cell counts measured by FACS, *Plac8* is ranked 1st, 97th and 22nd for the H1N1, pH1N1 and H5N1 cohorts, respectively, whose human protein interacts with PB1 [84]. *Nudc* is ranked 7th and 288th for the pH1N1 and H5N1 cohorts (12499th for the H1N1 cohort). The human NUDC protein interacts with M1, HA, and NA proteins [84]. Similar to the adjusted gene expression $\boldsymbol{g}$, we find *Sqstm1* and *Mapk7* among the top 10 genes. Furthermore, *Irg1* is ranked 4th and 34th for pH1N1 and H5N1 cohorts, respectively. This gene is considered significant through standard differential analysis while regarded as rather insignificant by gene expression adjusted by predictions from CIBERSORT with FDR values greater than 0.01 for all sample cohorts. This indicates that using adjusted gene expression values is better isolating gene expression associated directly with virus replication while using unadjusted expression is isolating immune response events induced by the infection while the ability of identifying disease-associated genes depends on the approach of adjustment.

When considering functional enrichment analysis of DE gene sets, unadjusted expression identified several key virus-infection-associated processes while adjusted gene expression filters out part of the annotations associated with inflammatory response. For example, significant genes (FDR $< 10^{-4}$ for at least two timepoints, see Materials and Methods) obtained from adjusted gene expression $\boldsymbol{g}$ for the pH1N1 cohort are enriched in antigen processing and presentation, 4-iron-4-sulfur cluster binding and small GTP-binding protein domain. The H1N1 and H5N1 cohorts are enriched in Ras-association, endoplasmic reticulum (ER), protein transport and proteasome. All three sample cohorts were enriched in cadherin binding involved in cell-cell adhesion, ribosomal protein. In contrast, significant

genes from standard DE analysis were associated with annotations related to response to virus and 2-5-oligoadenylate synthetase for all sample cohorts. Both the pH1N1 and H5N1 cohorts were highly enriched in chemotaxis, cytokine activity, cellular response to interferon-gamma and cell cycle. For significant genes calculated from adjusted gene expression $\boldsymbol{e}$, their enriched functional annotations were similar to those for standard DE. All sample cohorts were enriched in $2'$-$5'$-oligoadenylate synthase, innate immune response, and both pH1N1 and H5N1 cohorts are enriched in cellular response to interferon-gamma. The H1N1 cohort is also related to RNA binding and regulation of transcription, and the pH1N1 is associated with antigen processing and presentation as well as GTP binding.

3.4 Conclusion and Discussion

In this work, we applied four deconvolution algorithms to microarray data from mouse lung tissue infected by the influenza virus and tested their accuracy on predicting either absolute or relative changes in cell counts. Most algorithms predict well on several cell types for select sample cohorts except MLLSR, which is unsuccessful in the majority cohorts/cell types tested and has the lowest R^2 value. We utilized the estimated cell quantities to modify differential expression analysis and demonstrated that the adjusted gene expression largely improves the statistical significance of disease-associated genes and the efficiency of discovering key factors.

The first major caveat of the presented work is that the data is unpaired. Ideally, the gene expression and cell count data would come from the same animals. But in mice studies, tissue quantity is limited, and it is very common for pathology studies to have unpaired data (i.e., data from different animals) [29, 89, 90, 91]. We justify our approach here of evaluating the ability of the algorithms to predict the mean cell counts (or mean change in cell counts) as the animals used for lung gene expression and those for lung cell counts were infected with the same viruses, at the same initial loads, in the same laboratory and demonstrated the same symptoms and disease characteristics over time (e.g., weight loss, general lethargy and death due to infection were consistent [29]). Furthermore, the

mice having the same genetic background were housed in the same facilities and treated in the same manner during the experiments. Future work will focus on collecting paired gene expression and cell count data to provide a more thorough analysis of deconvolution algorithm's predictive performance. Yet, this work is still highly significant as whole genome gene expression analysis is a common and established tool to characterize disease-associated gene expression, and the findings summarized below strongly suggest that changes in tissue cellular composition must be addressed to improve prioritization of disease-associated gene candidates.

Based on the analysis of microarray data of influenza-infected mouse lungs, we find that the clustering-based algorithm, CTen, provides the most accurate cell count estimates. This improved predictive performance is likely due to the ability of CTen to infer across time points, whereas MLLSR, CIBERSORT and DCQ compute each time point independently. In addition, the regression-based tools analyzed here require much more reference data than CTen. These problems raise additional questions about the ability of select regression-based algorithms to infer accurately when applied to tissue samples, even though they demonstrated reasonable performance in the deconvolution of samples comprised of only 3 or 4 cell populations [67, 70]. Therefore, future research on cell count inference of time-course transcriptional profiles should consider either the more advanced clustering technologies (e.g., t-SNE [92]), or the models in the regression algorithms should be modified to account for time-course dynamics.

The regression-based algorithms can be utilized to infer cell quantities when applied to data from single time points and if the suitable reference data is available. This study finds that CIBERSORT provides more accurate predictions than MLLSR or DCQ for expression profiles sampled at a single time point. We also find that the inference accuracy of the tested algorithms is independent of the scale of the cell count data for CIBERSORT, MLLSR and CTen, but DCQ shows a significant association with the scale of the data having an R^2 value of 0.26 (as shown in Figure 29), suggesting that the accuracy of DCQ is biased to larger cell populations.

Finally, we demonstrate that adjusting gene expression due to changing cell populations within the tissue improves the identification of disease-associated genes. However, the

discovery of influenza virus-associated genes was improved at the cost of weakening the identification of biological functions. A possible explanation is that we utilized a strict cutoff of FDR values for all sample cohorts (FDR $< 10^{-4}$ for at least two timepoints) in order to limit the number of genes for functional annotation analysis. Another important reason may be that the adjustment due to changing cellular composition was moderately accurate since the estimated cell fractions have an overall R^2 value no larger than 0.4 when compared with measured cell counts. But despite the limitation, our study emphasizes that the adjustment of cell composition applied to transcriptomic data improves identification of meaningful genes which could be used as potential drug targets. Further improvement of deconvolution algorithms will greatly advance the systems biology and bioinformatics communities' ability to accurately model complex disease in tissue and improve the discovery of disease-associated genes.

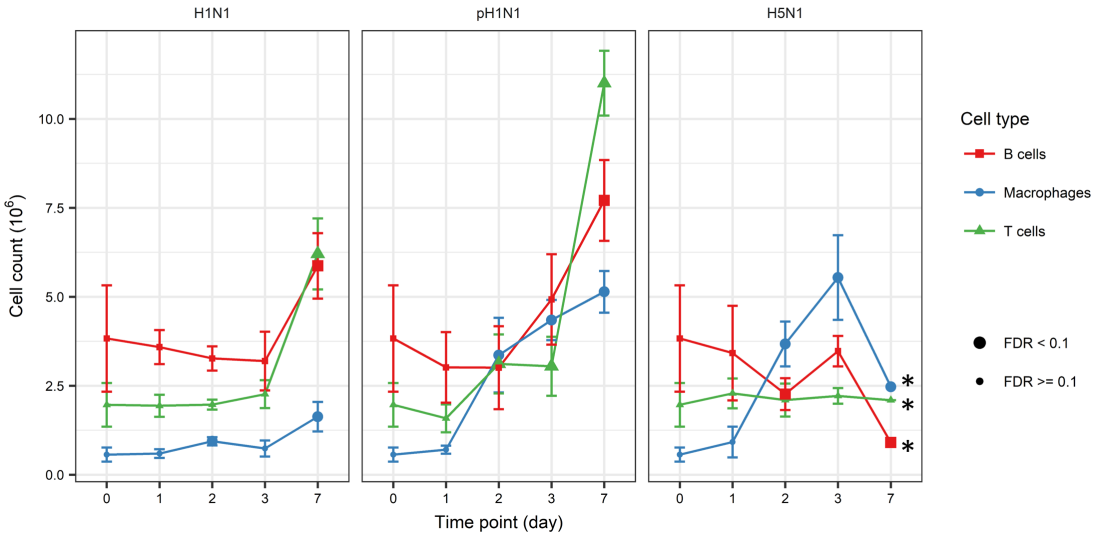
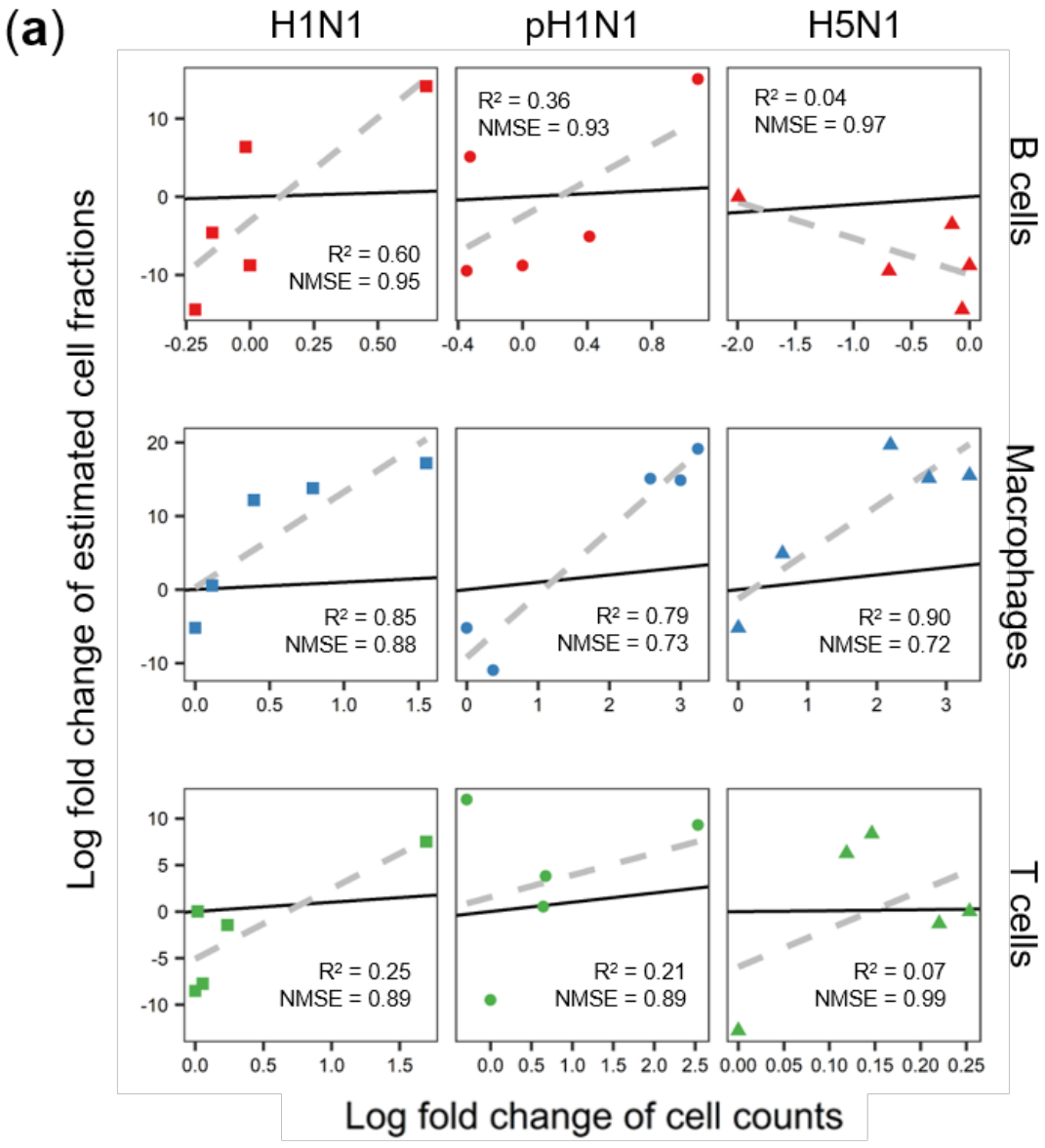


Figure 5: Cell counts of B cells, macrophages, and T cells in mouse lungs after infection by either H1N1, pH1N1 or H5N1 virus. Day 0 data are from uninfected, control animals. * Animals infected by H5N1 died before day 7.


(a)
H1N1
pH1N1
H5N1
B cells
Macrophages
T cells
Log fold change of estimated cell fractions
Log fold change of cell counts
R² = 0.60
NMSE = 0.95
R² = 0.36
NMSE = 0.93
R² = 0.04
NMSE = 0.97
R² = 0.85
NMSE = 0.88
R² = 0.79
NMSE = 0.73
R² = 0.90
NMSE = 0.72
R² = 0.25
NMSE = 0.89
R² = 0.21
NMSE = 0.89
R² = 0.07
NMSE = 0.99

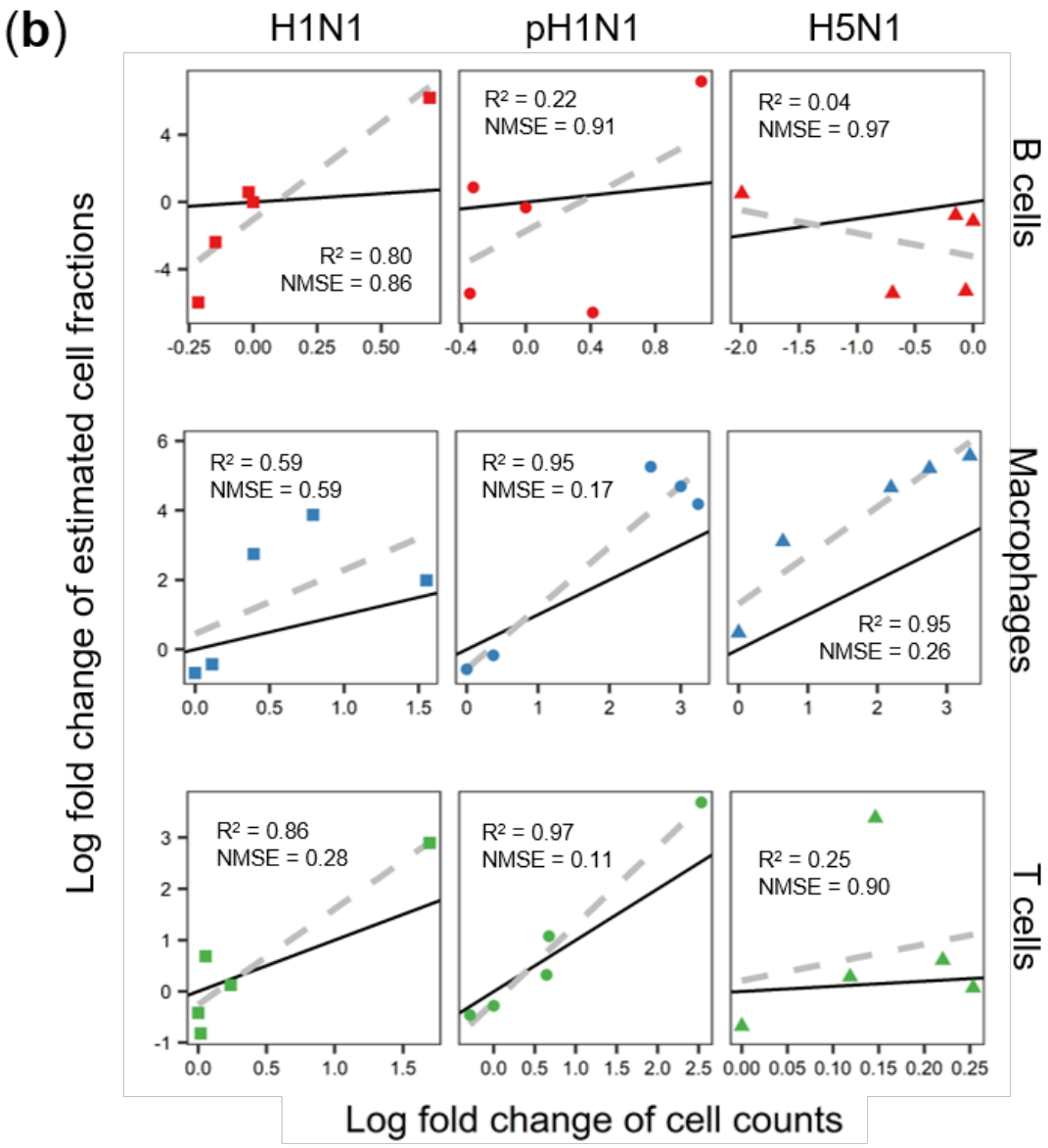
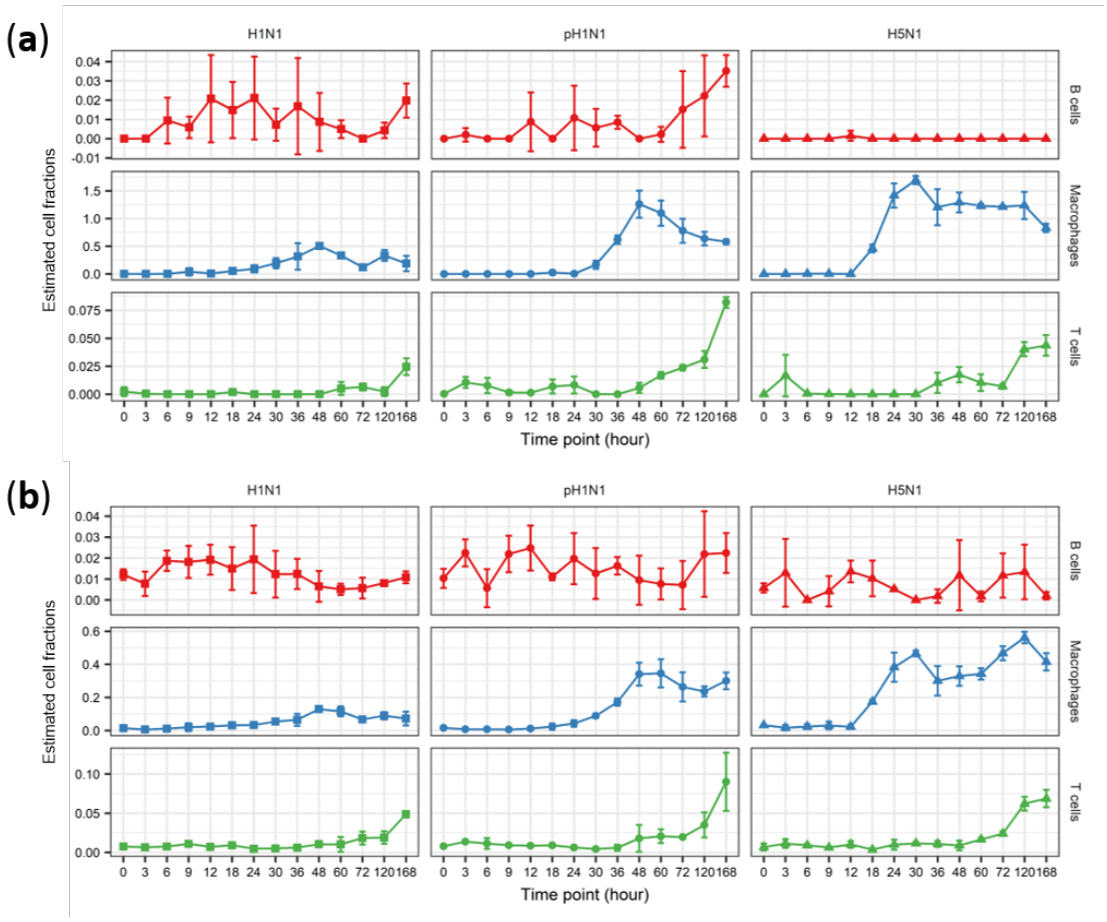


Figure 6: Log fold change (virus infection versus mock) of estimated cell fractions by modified linear least-squares regression (LLSR) (**a**) and cell-type identification by estimating relative subsets of RNA transcripts (CIBERSORT) (**b**) in comparison with log fold change of cell counts measured by fluorescence-activated cell sorting (FACS) for B cells, macrophages and T cells. Concordance is characterized by R^2 and normalized mean squared error (NMSE) values. The black line is $y = x$ while the grey dashed line is regression.


(a)
H1N1
pH1N1
H5N1
B cells
Macrophages
T cells
Estimated cell fractions
Time point (hour)
(b)
H1N1
pH1N1
H5N1
B cells
Macrophages
T cells
Estimated cell fractions
Time point (hour)

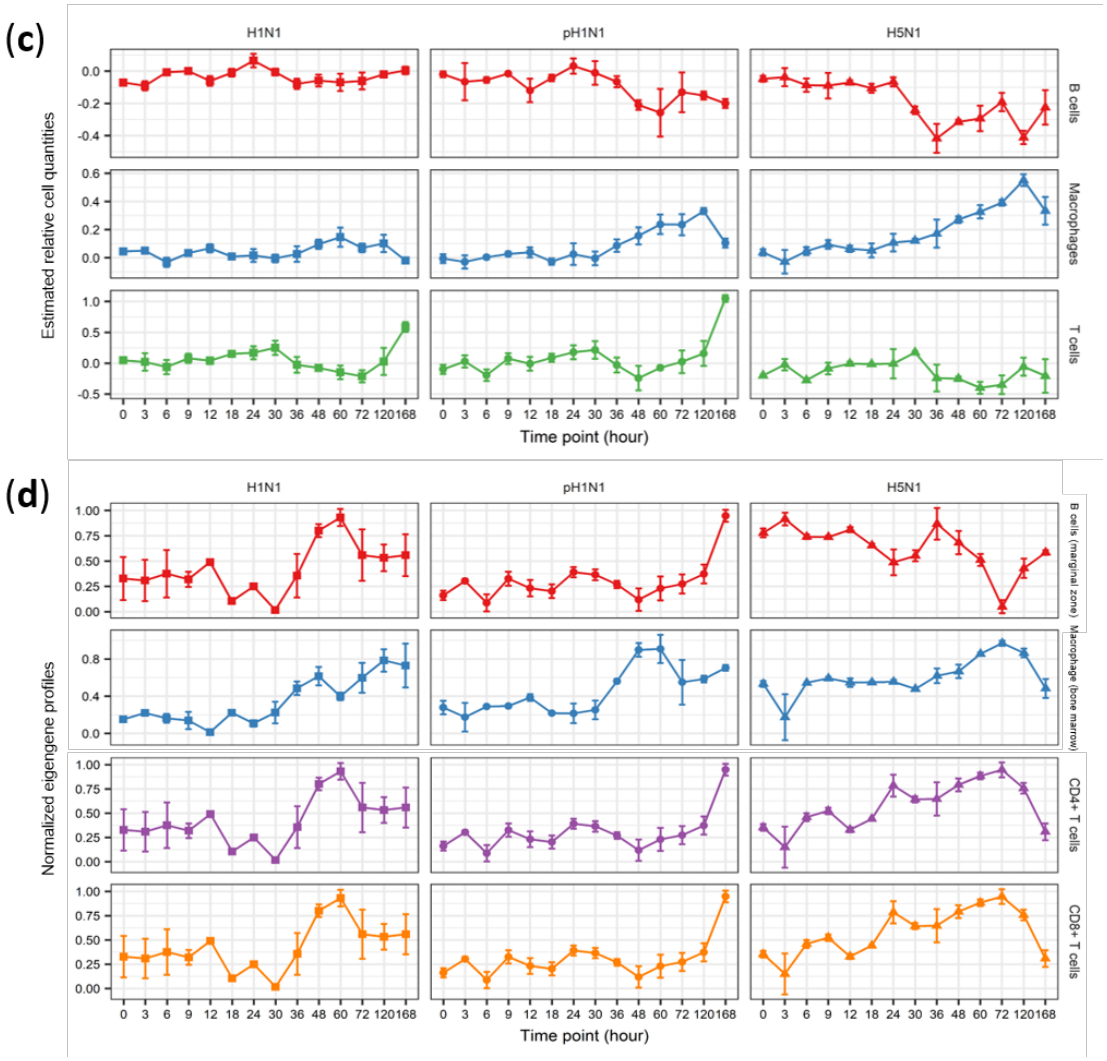


Figure 7: Estimated cell quantities of B cells, macrophages and T cells across time obtained using (**a**) MLLSR, (**b**) CIBERSORT, (**c**) digital cell quantification (DCQ), and (**d**) cell type enrichment (CTen). There were three samples per time point. Error bars depict the standard deviation of the estimate.

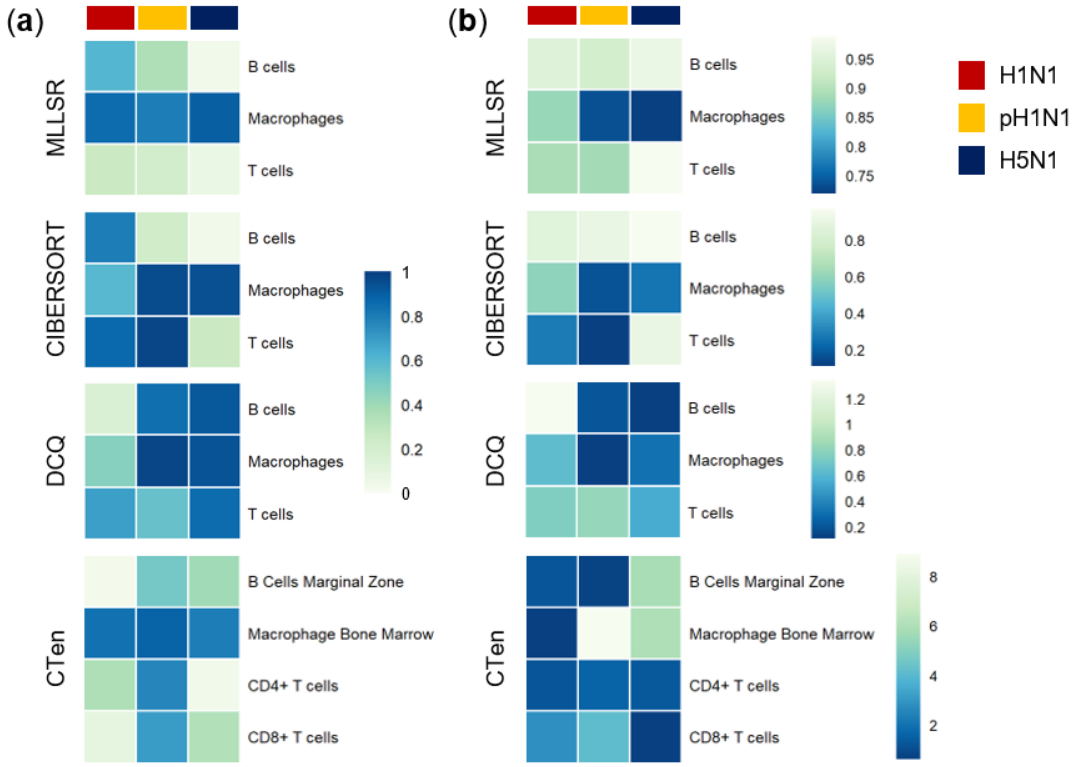


Figure 8: Two measurements for the accuracy of predicted cell quantities. Log fold changes of estimated cell fractions from MLLSR or CIBERSORT are compared with log fold changes of cell counts at the same time point per virus strain to compute R^2 (**a**) and NMSE (**b**). Similarly, R^2 (**a**) and NMSE (**b**) values for DCQ are calculated as estimated relative cell quantities versus change of normalized cell counts. R^2 (**a**) and NMSE (**b**) of CTen's predictions are calculated as eigengene profiles against log fold changes of cell counts.

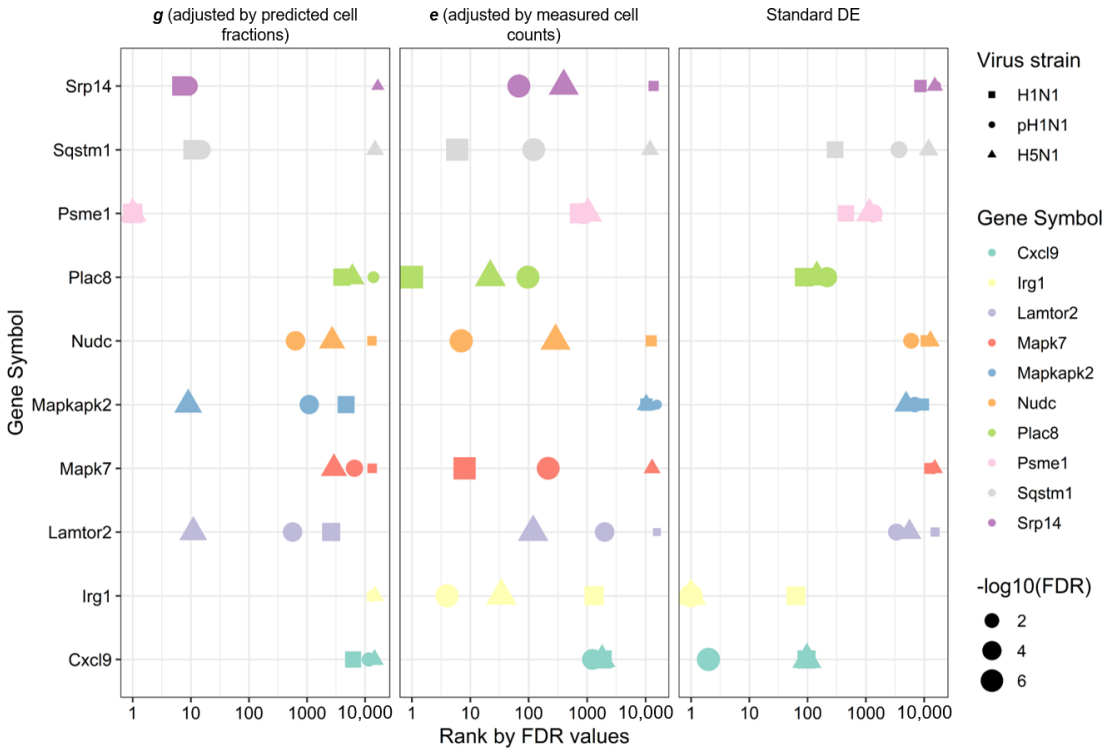


Figure 9: Ranking and false discovery rate (FDR) values of significant genes from calculating adjusted gene expressions and standard differential expression (DE) analysis.

4.0 Network Insights into Improving Drug Target Inference Algorithms

4.1 Introduction

The innovation of treatments for diseases remains a challenging task [93, 6, 94, 95, 96, 21]. The efficiency of pharmaceutical research and development (R&D), quantified by the number of new drugs per billion US dollars spent, dramatically declined from 1950 to 2010 [6]. A large group of drug candidates fail in clinical trials because they are not effective or safe in humans [6, 96, 97, 98, 12]. A major reason is that the systematic effects of drug candidates are not well studied or modelled in the drug discovery process, and a better understanding of their mechanisms of action (MoA) can help improve the efficiency of drug R&D [6, 99, 100, 101].

Two types of computational approaches have been reported to study MoAs of drugs by modeling high-throughput biological data: comparative analysis and network-based algorithms [102, 31]. Comparative analysis approaches, such as the Connectivity Map [103], have been used to predict molecular targets of drugs and assist in drug repurposing [103, 104, 105, 106, 107, 108]. They utilize expression profiles as drug signatures and compare with drugs having known targets, assuming that drugs with high similarities share the same targets. These approaches much rely on prior knowledge of drugs, thus have limitations in predicting de novo targets.

Network-based algorithms predict drug or disease targets by combining network information and transcriptomic data [31, 109, 30, 110, 111, 112, 113, 114]. Two recent representatives, DeMAND [30] and ProTINA [31], model the systemic dysregulation of regulatory network caused by a drug treatment, connecting molecular interactions with differential expression (DE). The regulatory network is generated by using protein-protein interactions (PPIs) and protein-gene interactions (PGIs) obtained from self-curated or public databases, such as STRING [115] and CellNet [116]. Similar to what has been reported by Noh et al [31], our preliminary research showed that ProTINA outperforms DeMAND when tested by the same gene expression and network datasets (p-value < 0.0002, Figure 30). Therefore, in this work, we focused on studying ProTINA's performance.

For ProTINA, a regulatory network, directing from proteins or transcription factors (TFs) to regulated genes, is generated from input PPIs and PGIs based on certain rules (Figure 10) [31]. The assumption is that the log fold change (LFC) of a gene is the linear combination of the LFCs of all proteins and TFs that regulate it. The weights are computed by linear regression methods and then integrated into a score for each regulator, a protein or a TF. Different from DeMAND, ProTINA may result in negative or positive scores, representing attenuation or enhancement, respectively. Regulators of larger magnitudes are more likely to be targets (red nodes in Figure 10). Showing promising results in predicting targets for *in vitro* datasets, DeMAND and ProTINA have provided a new direction in identifying drug targets and toxicity [31, 30].

However, as target inference algorithms become more complicated, it is unclear what roles gene expression and network data play. For predictions of perturbation patterns by biochemical models, network topology shows more impact on prediction accuracy than any other types of biological data, such as kinetic parameters [32]. Several studies have shown that proteins associated with disease and proteins that are drug targets have significantly different positions within biological networks [117, 118, 119, 120]. For target inference algorithms, it remains an open question as to which kind of biological data most affects the accuracy. Furthermore, algorithms can infer drug targets in a cell/tissue type-specific manner [31, 30, 114], and it is unknown how efficient or meaningful cell/tissue type-specific network data is for target inference. Answering these questions can provide us with insights into future algorithm improvement.

In this work, we evaluated the impact of gene expression and network data, using the human B cell microarray data from the DREAM challenge (referred to as DP14) as our benchmark dataset [121], and introduced a new algorithm to predict drug targets. Firstly, we found that ProTINA's scores are mostly determined by network data through permutation tests on gene expression. Secondly, we tested how the selection of networks affects prediction accuracy. Surprisingly, the effects of size or cell type are negligible. Next, our analysis suggested that network betweenness values can accurately predict drug targets. Different from network degrees, the accuracy is consistent regardless of the network size, and the performance is comparable with ProTINA.

Based on these findings, we proposed TREAP as an alternative approach to ProTINA, which combines the ranks obtained from betweenness values and adjusted p-values to predict drug targets. Here, we want to highlight the advantages of TREAP compared with other similar approaches. Firstly, TREAP has consistent performance on different datasets, and the median area under the receiver operating characteristics (AUROC) values are above 0.800 for all three datasets studied in this work. The median AUROC values are also higher than ProTINA and DeMAND for the same datasets. Secondly, TREAP is significantly more time-efficient than most other methods. Moreover, the simplicity and flexibility of the algorithm makes it more tractable to users who are not experts in systems and network biology. Thirdly, we chose betweenness values because our analysis showed that compared with degrees, they are less affected by the change in network size.

4.2 Methodology

4.2.1 Gene Expression Data Used in the Analysis

The microarray data of human Diffuse Large B-Cell Lymphoma (DLBCL) OCI-LY3 cell line treated with 14 different drugs under diverse doses at 3 time points, 6, 12 and 24 hours post treatment were obtained from the NCI-DREAM challenge drug synergy dataset, DP14 (GEO accession: GSE51068) [121]. Three samples treated with 'Aclacinomycin A' under a lower dose were dropped due to less significance. The microarray data of human liver cell line HepG2 treated with 62 genotoxic or non-genotoxic chemicals at 12, 24 and 48 hours post treatment were obtained from literature, referred to as HepG2 in this work (GEO accession: GSE28878) [122]. The microarray data of mouse pancreatic cells treated with 29 chromatin-targeting compounds were also obtained from GEO database, referred to as MP (GEO accession: GSE36379) [123]. For all three datasets, raw data were normalized using the RMA function from the "*affy*" R package [124]. The log2 fold change (LFC) values and Benjamini–Hochberg adjusted p-values (adjusted p-values) were calculated by the "*limma*" R package [72]. Probes were mapped to gene symbols by using the "*hgu219.db*" R package

[125] for human microarray data and "*moe430a.db*" [126] for mouse data. Those with the lowest average BH-adjusted p-value across all samples were chosen when multiple probes were mapped to the same gene.

4.2.2 Networks Used in the Analysis and Calculation of Topological Features

Human or mouse PPIs and their associated confidence scores were obtained from the STRING database [115]. Interactions with experimental proof or from curated databases (the channels of 'experiments' and 'databases') were extracted. Interactions transferred from other species or duplicated entries were excluded. Subnetworks were obtained by applying thresholds ranging from 0.4 to 0.9 to the PPI network, referred to as PPI04, PPI05, PPI06, PPI07, PPI08 and PPI09, respectively in this paper.

Human PGIs and their confidence scores were obtained from the Regulatory Circuits, a database of predicted, cell/tissue type-specific PGIs [127]. PGI networks of 8 different cell/tissue types were studied in this analysis: 'lymphocytes of B lineage', 'lymphocytes', 'lymphoma', 'myeloid leukocytes', 'lung', 'heart', 'epithelial cells' and 'hepatocellular carcinoma cell line'. The network of 'lymphocytes of B lineage' were predicted by samples including those from DLBCL, the same cell line with DP14 [127], thus was chosen as a reference for analysis of DP14. PGI subnetworks for each cell/tissue type, namely PGI05, PGI10, PGI15 and PGI20, were obtained by thresholds ranging from 0.05 to 0.20, respectively. Mouse PGIs were compiled from two manually curated databases of transcriptional regulatory networks: TRRUST (version 2) [128] and RegNetwork [129]. These interactions are not cell/tissue type-specific, and no threshold was applied to them prior to analysis of MP.

Degree or betweenness values were calculated by the "*igraph*" R package [130]. PPIs or the combination of PPIs and PGIs (PPI+PGI) were treated as undirected graphs, while PGIs were treated as directed graphs.

4.2.3 Reference Drug Targets

The reference targets of each chemical were extracted from STITCH database (version 5.0) [131, 132] for analyses of DP14, HepG2 and MP. Only targets with experimental proof or from curated databases were collected.

4.2.4 Prediction of Drug Targets by ProTINA

LFC values, PPI and PGI subnetworks were analyzed by "*protina*" R package [31]. Slope matrices of each time point were calculated following the user manual. For samples with only two timepoints, control samples served as 0hr post treatment to calculate associated slope matrices. Samples from different doses for the same drug were treated as separate groups.

4.2.5 Prediction of Drug Targets by TREAP

For target prediction by TREAP, the assumption is that genes with high betweenness values or low adjusted p-values are more likely to be drug targets. There are three steps for this algorithm. The first step is to calculate PPI+PGI betweenness values and obtain adjusted p-values from DE analysis. For gene expression profiles with multiple timepoints, DE analysis can be performed per time point or across all timepoints. The second step is to calculate the ranks of genes by sorting betweenness values and adjusted p-values, respectively. Genes with high betweenness values or low adjusted p-values are scored with high ranks. The third step is to generate final scores by summing up the ranks from both metrics for each gene. Genes with higher scores are more likely to be targets for associated drugs.

Adjusted p-values and PPI+PGI betweenness values were calculated as explained in the former sections. Ranks of genes were obtained by sorting betweenness values and adjusted p-values, respectively, and genes with the same betweenness or adjusted p-value shared the same rank. Final scores were calculated by summing up the ranks from both metrics for each gene. In this work, all analyses on TREAP used 0.9 as the threshold for human or mouse PPIs and 0.20 for human PGIs. No threshold was applied to mouse PGIs.

4.2.6 Calculation and Comparison of AUROC Values

Area under the receiving operator characteristics (AUROC) values in this paper were calculated by comparing scored proteins with reference drug targets through the "*pROC*" R package [133]. As ProTINA scores can be positive or negative, the absolute scores were used to calculate AUROC. The median AUROC across all drugs in each dataset was calculated to represent accuracy of a whole test. For drugs having more than one doses, the AUROC values of low doses were excluded. In terms of topological features, degree or betweenness values were directly used to calculate AUROC values without pre-processing. TREAP scores were directly used for calculation of AUROC without preprocessing. Difference in any pair of chosen tests were computed by performing pairwise t-test between their AUROC values. A p-value less than 0.05 were regarded as significantly different.

4.2.7 Permutation Tests on Gene Expression

The null hypothesis for permutation tests in this work is that the median AUROCs of randomized gene expression are smaller than that of nonrandomized gene expression, and the p-values were calculated accordingly. For ProTINA, gene labels for DP14 that refer to the rows of LFC and associated slope matrices were randomly shuffled for 1000 times. Randomized data were applied to ProTINA under the same network setup, PPI09 and PGI20. For TREAP, gene labels of adjusted p-values for each dataset, namely DP14, HepG2 and MP were randomly shuffled for 1000 times, respectively, and drug targets were predicted using PPI09 and PGI20 (for MP no threshold was applied). AUROC and median values were calculated as explained in the former section.

4.3 Results

4.3.1 Permutation Tests Show that ProTINA is Predominantly Determined by Network Data

To understand how much network or gene expression data contribute to ProTINA's accuracy, we performed 1000 permutation tests by randomizing the LFC gene expression values (Materials and Methods). To shorten the computation time on ProTINA, the smallest PPI and PGI subnetworks (PPI09 and PGI20) were chosen for this analysis (discussed more in the following section). For prediction scores obtained from ProTINA, the area under the receiver operating characteristics (AUROC) values were calculated per drug, and the median AUROC across all drugs is used as a metric for each test.

The median AUROC obtained by nonrandomized LFCs is 0.799. As shown in Figure 11, 220 of 1000 permutation tests have higher median AUROC values than that (one-tailed p-value = 0.221). Most tests have similar accuracy to the original test. This indicates that randomizing LFCs does not diminish ProTINA's accuracy significantly, therefore, network data determines most of ProTINA's performance.

4.3.2 Selection of Networks Has Limited Effects in the Prediction Accuracy of ProTINA

Next, we studied how the selection of PPIs and PGIs affects target inference accuracy. PPI and PGI subnetworks of different sizes or cell/tissue types were tested using the same gene expression data from DP14 (Materials and Methods). Similar to permutation tests, the median AUROC represents the accuracy for each PPI-PGI combination.

In total, 24 PPI-PGI combinations of different sizes were tested on ProTINA. As shown in Figure 12a, ProTINA favors small PPI and large PGI subnetworks. The combination of PPI09 and PGI05 shows the highest accuracy, and its median AUROC is 0.821. As the threshold increases from 0.4 to 0.9, the number of interactions in PPI subnetwork ranges from 380375 to 281357 (Figure 31), and the median AUROC increases for most tests. For example, the median AUROC increases from 0.785 (PPI04) to 0.811 (PPI09) for analyses

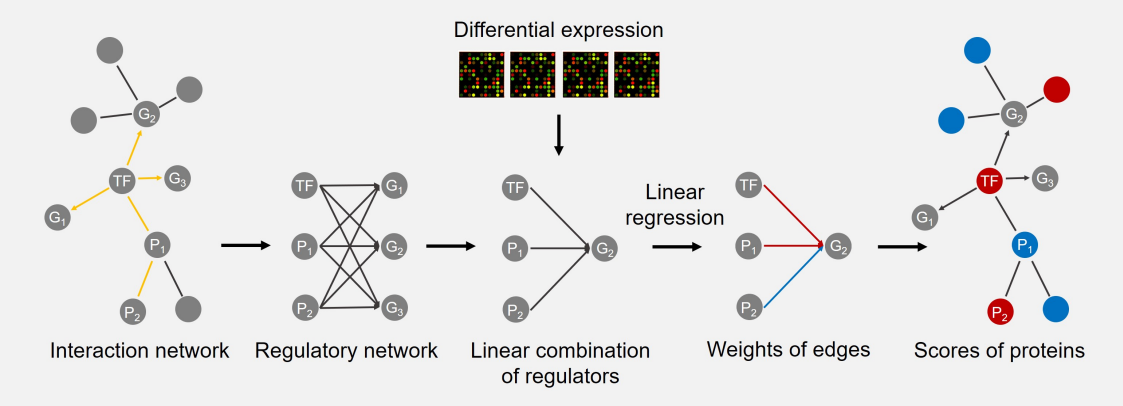


Figure 10: An overview of ProTINA algorithm. Each node refers to a transcription factor (TF), a non-TF protein (P) or a gene (G). Arrows present the directions of interactions or edges. The significance of an edge or protein (including TFs) is color coded, where red refers to high significance while blue refers to low significance.

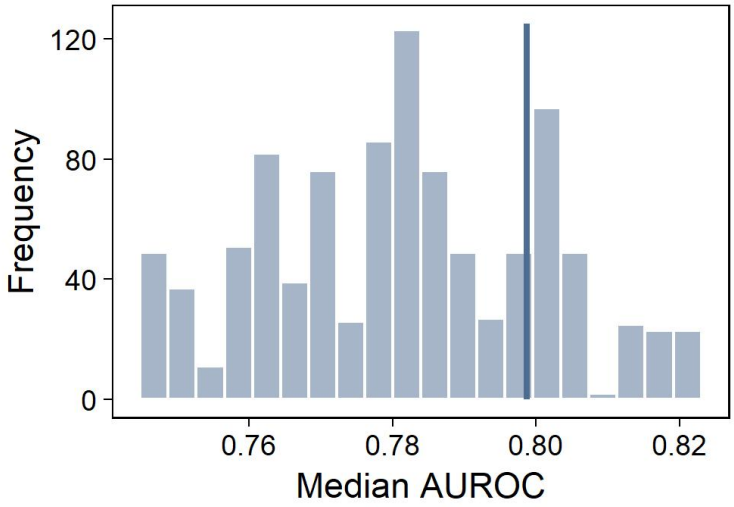


Figure 11: 1000 Permutation tests were performed by randomizing the gene expression and calculating the median AUROCs. The blue vertical line refers to the median AUROC obtained by nonrandomized gene expression.

under PGI10. But there are PPI subnetworks that do not follow this trend. Under PGI05, PPI05 shows lower median AUROC than PPI04 (0.784 and 0.796, respectively). For PGI subnetworks, as the number of interactions range from 123394 to 5932 (Figure 31), the median AUROC shows an opposite trend to that of PPI subnetworks (Figure 12a). An example is that the median AUROC decreases from 0.821 (PGI05) to 0.785 (PGI20) when using PPI09. Most tests show a consistent trend except for those using PPI05. A possible reason is that new proteins and associated interactions are included in the network as the threshold for PGIs changes from 0.05 to 0.10. So that the network topology is changed, and predictions from ProTINA are affected accordingly.

However, median AUROC values vary in a small range while the size of PPI or PGI subnetworks changes significantly (Figure 12a, 12b and Figure 32). The highest and lowest median AUROC values for ProTINA are 0.821 and 0.753, with a difference less than 0.1 (Figure 12b). In addition, most of these differences are insignificant. When comparing with the test using PPI09 and PGI20, none of the 24 tests are significant in AUROC values (p-value > 0.05 for all tests), although the combination of PPI04 and PGI15 has resulted in a pairwise p-value of 0.057. Furthermore, we studied the effects on standard deviations (SDs) of AUROC values across 12 drugs for each test. All of them maintain at a low level below 0.13. In summary, we conclude that the size of networks has limited effects on the prediction accuracy of ProTINA, while small PPI and large PGI networks tend to improve the accuracy.

To analyze the performance of cell/tissue type-specific networks, 28 tests using PGIs from 7 cell/tissue types were performed on ProTINA using the same PPI subnetwork, PPI09. Most tests counterintuitively show similar median AUROC values regardless of cell/tissue types (Figure 12c, Figure 33). In theory, the PGI subnetworks for immune cells should have higher accuracy than non-immune cell types, and those for 'lymphocytes of B cell lineage' should outperform other immune cells. This is because that samples from DLBCL, the same cell line with DP14, were used to predict the interactions for 'lymphocytes of B lineage' [127] (Materials and Methods). However, using the AUROC values from PGI20 for 'lymphocytes of B cell lineage' as a reference, no other cell/tissue types are significantly different (pairwise p-values > 0.05 for all) under the same network setup. In conclusion, we have shown that

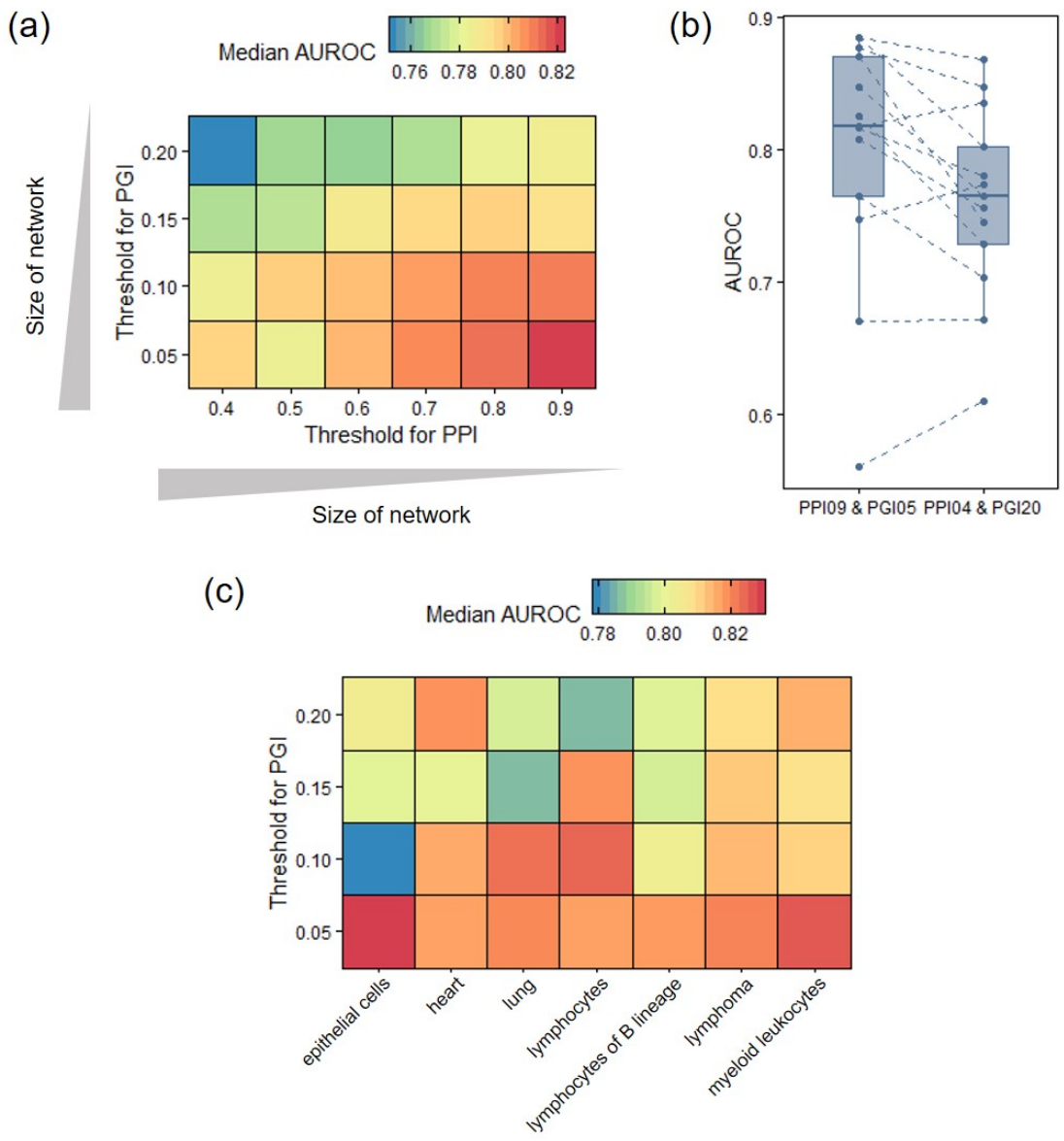


Figure 12: Prediction accuracy of ProTINA using networks of different sizes or cell/tissue types. (**a**) PPI or PGI subnetworks of different sizes were tested on ProTINA to predict targets for DP14. The axes refer to the confidence thresholds for PPI (x axis) and PGI (y axis) subnetworks, and the median AUROC values are the metric for prediction accuracy. Among all PPI-PGI subnetwork combinations, PPI09-PPI05 and PPI04-PPI20 have the highest and lowest accuracy in terms of median AUROC values, respectively. Panel (**b**) shows the boxplot of these two groups. Each dot represents the AUROC of a drug. (**c**) PGI subnetworks of 7 cell/tissue types were applied to ProTINA for target prediction.

the selection of PPIs or PGIs in terms of either the size or cell/tissue type is not the key factor to prediction accuracy of ProTINA.

4.3.3 Topological Features Have Similar Prediction Accuracy to ProTINA, and the Protein Betweenness Outperforms Degree

Our findings have shown that ProTINA depends on network topology more than gene expression, and that it has consistent performance regardless of the network size or the cell/tissue type the network represents. These suggest that ProTINA is probably determined by some network topological feature that remains relatively stable across different PPI or PGI subnetworks, such as protein degree or betweenness. The degree of a protein is the number of proteins/genes with which it interacts, while the betweenness is a measure of bottleneckedness, e.g. the amount of information flowing through the proteins that connect the rest of the network. Analyses of these features and their effects on drug target prediction may provide meaningful insights on improving prediction accuracy.

To test our hypotheses, we studied degree and betweenness values for PPIs, PGIs and PPI-PGI combinations (referred to as PPI+PGI in the following text). Firstly, for PPIs, we compared scores obtained from ProTINA (using PPI09 and PGI20) with their associated protein degrees or betweenness values in PPI09 for each drug. The majority of the drugs show a weak but evident correlation between absolute ProTINA scores and protein degrees, however, the correlation for betweenness is much lower (Table 3). For instance, the correlation coefficient is 0.211 for 'Rapamycin' (Figure 13a), while the correlation of betweenness values is smaller than that of degrees, which is 0.085 for the same drug (Figure 13b). Notice that a large portion of the top 100 proteins scored by ProTINA (red points in Figure 13a, b) lie in the group of high degree or betweenness values.

We next tried to predict drug targets by using PPI degree or betweenness values without considering gene expression or PGIs. The assumption is that proteins with higher degree or betweenness values are more likely to be targets. As shown in Figure 13c, the median AUROC values for PPI degree or betweenness values are close to those for ProTINA. What's more, betweenness values perform better than degrees. The highest and lowest median AUROC

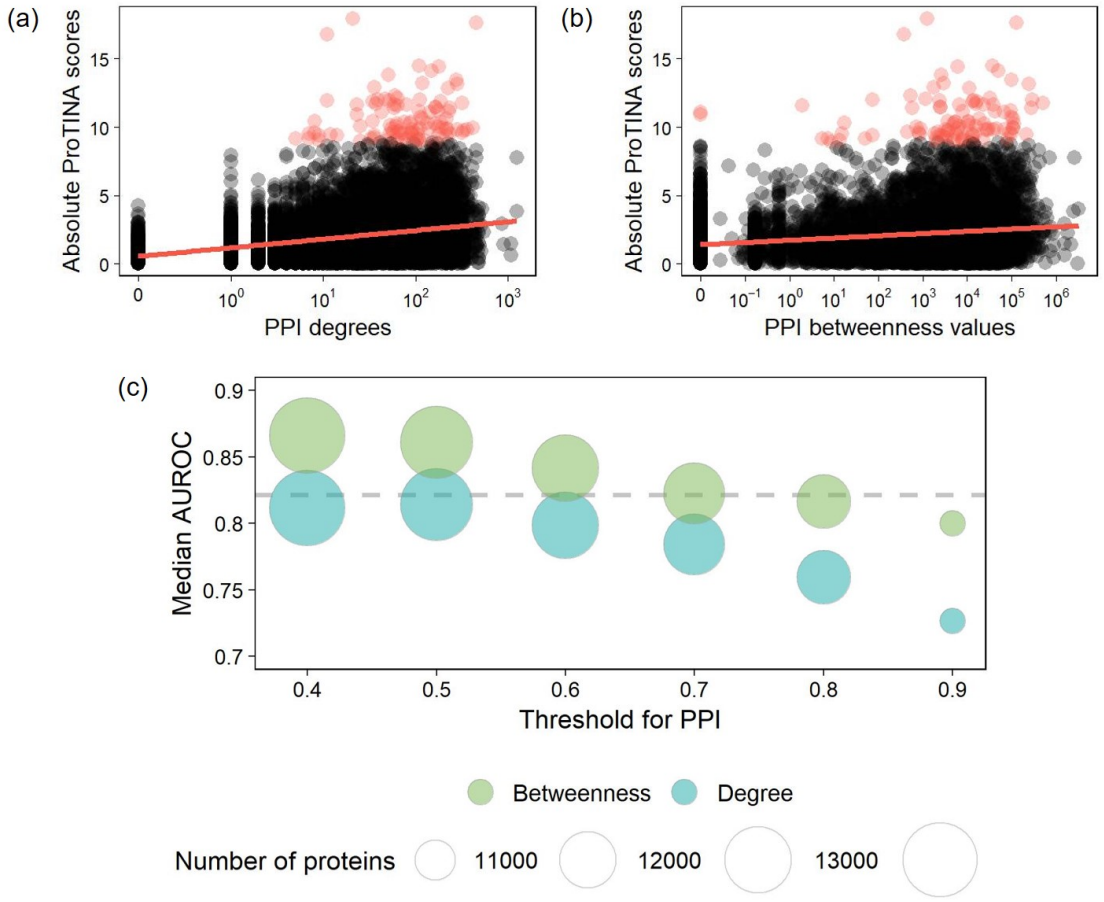


Figure 13: (**a**) Degree or (**b**) betweenness values of proteins in PPIs were compared with associated ProTINA scores for 'Rapamycin'. The correlation coefficient is 0.211 for absolute ProTINA scores and degrees, and that for betweenness values is 0.085. Red points refer to the top 100 proteins scored by ProTINA. (**c**) The degree and betweenness values were used to predict drug targets assuming higher scores are more likely to be targets. Each point shows the median AUROC value and the number of proteins under a PPI threshold. For reference, the grey dashed line refers to the highest median AUROC achieved by ProTINA, which was obtained by using PPI09 and PGI05.

values for degrees are 0.814 and 0.727 (Figure 13c), and those for betweenness values are 0.866 and 0.800, even higher than associated median AUROC values for ProTINA under the same network setup. As the size of PPI subnetworks shrinks, the median AUROCs for degrees decreases accordingly (Figure 13c), with a correlation coefficient value of -0.950 between the medians and thresholds. While the decrease of network size also diminishes the accuracy of betweenness, the median AUROCs remain higher than those of degrees and decrease relatively slower.

Secondly, the degree and betweenness values of PGI subnetworks were also compared with associated ProTINA scores, however, there are no clear trends between them (Table 3). In addition, drug target prediction based on PGI degree or betweenness values are not comparable with that by PPI topological features (Figure 34). This might be related to the limited amount of PGI interactions.

Lastly, we calculated topological features for PPI+PGI and compared them with ProTINA scores. As expected, they show the same trend with PPIs (Table 3). The correlation coefficient between degrees and ProTINA scores is 0.208 for 'Rapamycin', while that for betweenness values is 0.079 (Figure 14a, b).

Predicting drug targets by PPI+PGI degree or betweenness values results in higher median AUROC values than those for PPIs. In addition, for all thresholds of PPIs or PGIs applied to this analysis, betweenness values outperform degrees. The accuracy for PPI+PGI degrees ranges from 0.833 to 0.733 in terms of median AUROC values, and that for betweenness values ranges from 0.878 to 0.782 (Figure 14c). As the size of PPI or PGI subnetworks decreases, the median AUROC values for PPI+PGI degrees also decreases. PPI+PGI betweenness values have the same behavior as the size of PPIs changes, while for PGIs the trend is less evident. PGI10 has the best performance in parallel comparisons. In summary, betweenness values well predict drug targets and show even higher accuracy than ProTINA.

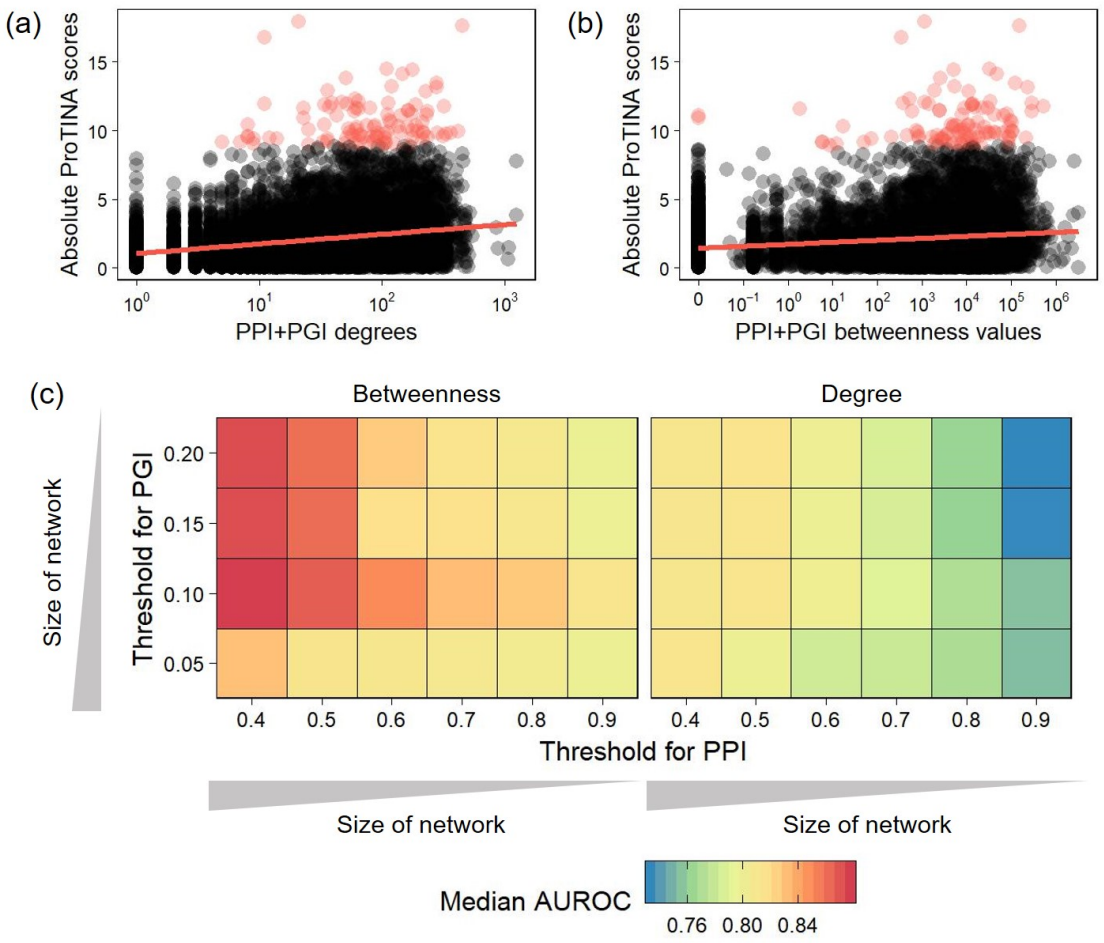


Figure 14: (**a**) PPI+PGI degree or (**b**) betweenness values were compared with associated ProTINA scores for 'Rapamycin'. The correlation coefficient for absolute ProTINA scores versus degrees and absolute ProTINA scores versus betweenness values is 0.208 and 0.079, respectively. Red points refer to the top 100 proteins scored by ProTINA. (**c**) The degree and betweenness values were used as measures to predict drug targets, and the median AUROC values were calculated for each prediction. The axes refer to the confidence thresholds for PPI (x axis) and PGI (y axis) subnetworks.

4.3.4 Missing Information in Network Topology Can be Covered by Differential Expression

Topological features, degree or betweenness values, have shown high prediction accuracy without taking gene expression into account. Our permutation tests have also indicated that network data has more effects in ProTINA's performance than gene expression data. What's more, ProTINA has much better performance than differential expression (DE) analysis on drug target prediction according to prior research [31]. All of the above has raised a question about whether gene expression data can help to predict drug targets. To address this concern, we compared DE analysis (adjusted p-values, Materials and Methods) with two other target prediction methods in terms of their performance on each drug: PPI+PGI betweenness and ProTINA.

We calculated AUROC values for all three methods using the same network setup, PPI09 and PGI20 (Figure 15). For most drugs, such as 'Mitomycin C' or 'Cycloheximide', PPI+PGI betweenness and ProTINA have close AUROC values, and they outperform DE analysis. Consistent behaviors between ProTINA and PPI+PGI further indicates the impact of network topology on ProTINA's accuracy. In contrast to these drugs, DE has much higher prediction accuracy than PPI+PGI or ProTINA for 'Monastrol' (the AUROCs are 0.998, 0.771 and 0.555, respectively). This means that DE analysis of gene expression data can capture information missing in network topology, and that it is necessary to include gene expression data for drug target inference and improvement of accuracy.

4.3.5 A Novel Algorithm that Combines Network Topology and DE Analysis for Target Inference

To better combine network topology and DE analysis and improve inference accuracy, we suggest TREAP (target inference by ranking betweenness values and adjusted p-values) to predict drug targets (https://github.com/ImmuSystems-Lab/TREAP). There are three steps for this algorithm. The first step is to calculate PPI+PGI betweenness values and obtain adjusted p-values from DE analysis. For gene expression profiles with multiple timepoints, DE analysis can be performed per time point or across all timepoints. The second

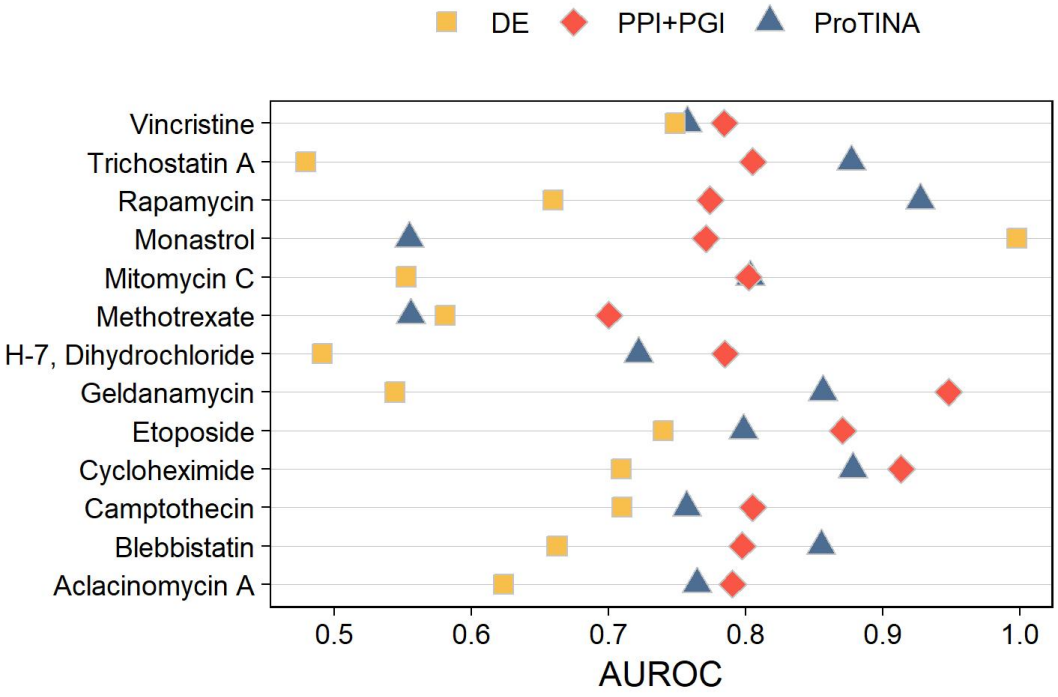


Figure 15: AUROC values of each drug obtained from three different methods: differential expression (DE) analysis by adjusted p-values, betweenness values from the combination of PPI09 and PGI20 (PPI+PGI) and ProTINA analysis by PPI09 and PGI20.

step is to calculate the ranks of genes by sorting betweenness values and adjusted p-values, respectively. Genes with high betweenness values or low adjusted p-values are scored with high ranks. The third step is to generate final scores by summing up the ranks from both metrics for each gene. Genes with higher scores are more likely to be targets for associated drugs.

TREAP was tested by three different gene expression profiles: (i) DP14 [121], (ii) human HepG2 cells treated with genotoxic or non-genotoxic chemicals, referred to as HepG2 in this work [122] and (iii) mouse pancreatic cell lines treated with chromatin-targeting compounds, referred to as MP [123]. Human and mouse PPIs were obtained from STRING [115], and PGIs were obtained from Regulatory Circuits [127], TRRUST (version 2) [128] and RegNetwork [129]. AUROCs were calculated by comparing scored genes with known targets for each test as a measurement of accuracy.

TREAP shows stable performance and maintains high accuracy for all datasets tested in this study (median AUROCs > 0.800, Figure 16). While TREAP takes significantly less computation time than ProTINA, it has higher median AUROCs when compared with ProTINA under the same dataset (Figure 16). For DP14, the median AUROC of TREAP is 0.850, higher than that of ProTINA, 0.799 (p-value $= 0.11$). Notice that it is also higher than using PPI+PGI betweenness values alone, which is 0.798. TREAP significantly outperforms ProTINA in HepG2. The median AUROC is significantly improved from 0.739 to 0.801, with a p-value of 0.0002. For MP, TREAP and ProTINA have close median AUROCs as 0.806 and 0.799, respectively (p-value $= 0.39$). By integrating betweenness values and adjusted p-values to represent both network topology and DE analysis, TREAP is comparable with and sometimes better than ProTINA in accuracy for all datasets analyzed in this work. In addition, we performed 1000 permutation tests on TREAP by randomizing the adjusted p-values for each dataset. Different from ProTINA, TREAP is significant when compared with permutation tests on the adjusted p-values from DP14, with the one-tailed p-value as 0.007 (Figure 35). For HepG2, the one-tailed p-value is 0.058, while for MP, TREAP is less significant and shows a one-tailed p-value as 0.314.

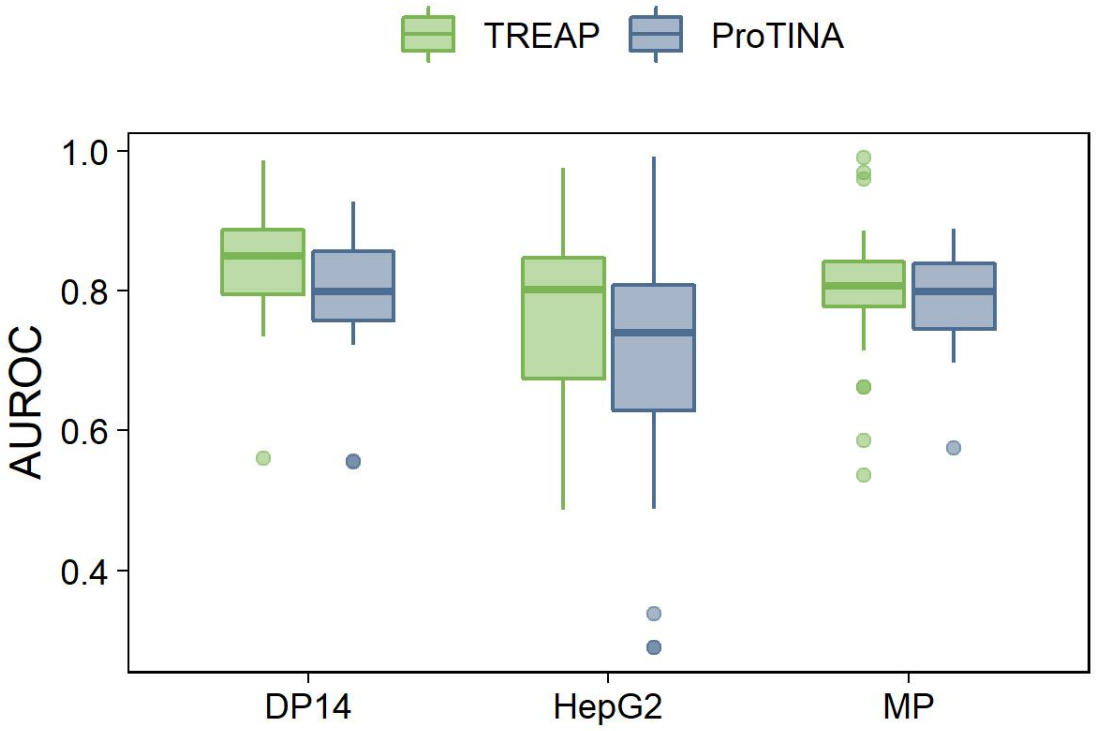


Figure 16: AUROC values of TREAP and ProTINA predictions for different gene expression profiles: human lymphoma cells (DP14), human liver cancer cells (HepG2) and mouse pancreatic cells (MP). (p-values = 0.11, 0.0002 and 0.39, respectively)

4.4 Conclusion and Discussion

Our analyses have shown that, even though ProTINA requires both gene expression and network data for inputs, network data predominantly determines accuracy of drug target inference. What's more, the cell/tissue type or size of a network has limited impact on ProTINA's accuracy, while topology, especially the degree value, affects the performance of ProTINA more.

However, ProTINA has two limitations due to the reliance on network topology alone and the connection with protein degrees. First, PPIs, a major part of network data, have a known bias toward protein abundance [119, 134]. It has been reported that interactions obtained from high-throughput experiments have a correlation between the protein degree and abundance. Second, we have shown that differential analysis of gene expression data can uncover meaningful information missing in network topology.

To address these two limitations, we suggested a new algorithm, TREAP, which combines protein betweenness values and adjusted p-values, representing information from both sources of network topology and DE analysis, to predict drug targets. We chose betweenness values because they are less sensitive to network sizes and more accurate than degrees in target prediction. TREAP shows more consistent performance than ProTINA when tested by different gene expression profiles and maintains a median AUROC above 0.800. In addition, TREAP takes significantly less computation time than ProTINA, and its simplicity makes it more tractable to users who are not experts in systems and network biology. It is also flexible in dealing with samples of limited or multiple timepoints as adjusted p-values can be calculated per timepoint or across all timepoints based on user's needs. However, ProTINA needs at least two timepoints to fully take advantage of the algorithm [31]. Currently, betweenness values and adjusted p-values are weighted equally for TREAP. Future work should focus on better balancing both types of data and trying other scoring methods to improve prediction accuracy.

TREAP is presented here as an alternative approach to ProTINA, but it is worth emphasizing advantages specific to each algorithm. As stated above, TREAP is significantly faster, and the algorithm is not complex, enabling users from several branches of research to

access the tool and understand the findings. ProTINA, however, is a mechanistically derived algorithm, which allows users with expertise in computational biology to dive deeper into the possible mechanisms of a drug's activity. The accuracy of their predictions is similar when measured using the AUROC, but the permutation tests presented here suggest that TREAP is more likely to use drug-specific gene expression to make a more accurate prediction.

TREAP and its derivatives have potential in a variety of applications for drug innovation. First, it can assist in selection of drug candidates and serve as a preliminary test of the efficacy or safety by connecting with databases for functional annotations, e.g. Gene Ontology [135, 136]. Studying predicted targets can help exclude poorly targeting drug candidates or those causing severe damage to biological systems. Second, the algorithm can be applied to drug repurposing by exploring published datasets characterizing drug treatments, assuming that a pair of drugs sharing the same group of predicted targets can be used to treat the same disease. Last but not the least, the algorithm can help to discover disease mechanisms [31]. Similar to drug treatments, diseases can also be treated as a type of perturbation to the biological system of interest. Predicting disease targets may assist in identifying key components of disease mechanisms and pathology, which is crucial for innovations in disease treatment [99].

5.0 Conclusions

This dissertation describes three projects involving analysis and interpretation of big biological data. These projects have demonstrated that computational approaches can help to exploit those data and better understand drug or disease mechanisms at the systems level for drug innovation and development. In the first project, we performed comparative analysis of RNA-Seq data from mice treated with different adjuvants and identified the mechanisms supporting adjuvant activity through collaboration with RIKEN institute in Japan. This training project shows that analysis of RNA-Seq data can provide insights for development of vaccine adjuvant. For the second project, we predicted immune cell dynamics by linear regression-based or statistical algorithms and suggested a new approach that can adjust gene expression to improve the discovery of key disease-associated genes. In this project, gene expression of a biological system is connected with changes in cell populations through computational tools. For the third project, it is shown that the network topology predominantly defines the accuracy of a major drug target inference algorithm, and we proposed a novel algorithm named TREAP, which integrates network betweenness and differential gene expression to accurately predict drug targets in a time-efficient manner. It is shown that computational approaches can connect molecular interaction data with gene expression to explore drug mechanisms. Future work should focus on better balancing both types of data and trying more sophisticated methods to improve the prediction accuracy of TREAP.

At the time of writing, three peer-reviewed articles associated with this work have been published. The work in predicting immune cell dynamics and key disease-associated genes is published in *IFAC* [1] and *Processes* [2]. Part of the adjuvant project is recently accepted by *International Immunology* [3], and one more article by our collaborators is expected. For the work in drug target inference, the R package of TREAP is now available for public use on GitHub, and the associated article is under review [4].

Appendix A

Supplementary Materials for Predicting Host Immune Cell Dynamics and Key Disease-Associated Genes

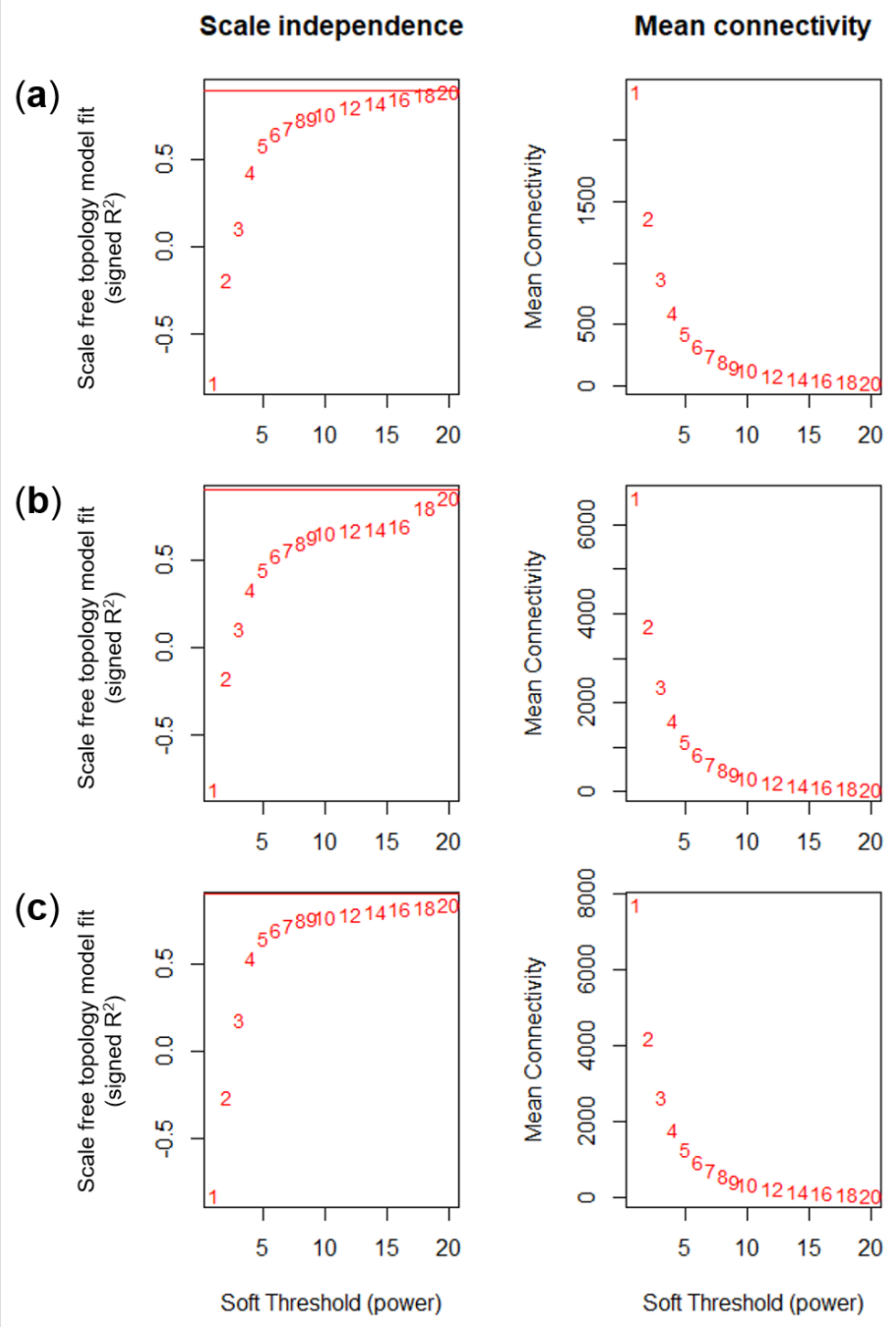


Figure 17: Scale-free topology fit using the WGCNA package for sample cohorts of (a) H1N1, (b) pH1N1 and (c) H5N1.

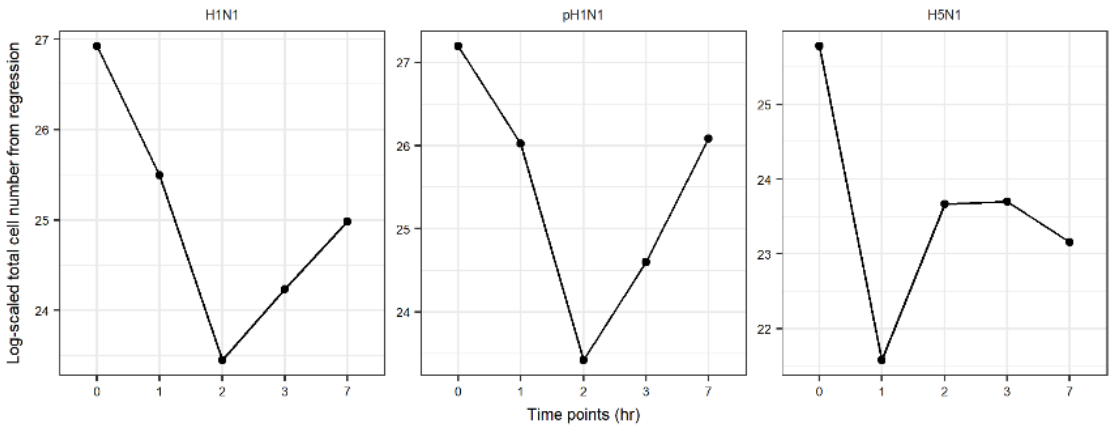


Figure 18: The estimated total number of cells in influenza virus-infected samples does not change significantly in the 7-day time course. Since the fraction of a cell type is equal to the number of this cell type divided by the total number of all cells in a sample, we have the following equation to connect total cell numbers with cell fractions: $c_t x_t = f_t$, where c_t is a vector of averaged cell fractions for n kinds of cells at time t predicted by CIBERSORT, f_t is a vector with the same length which represents averaged cell counts at time t measured by FACS, and the total number of cells at time t is represented as x_t. The values of x_t at all time points are estimated using linear regression (by the R function "lm"). We observe that these x_t values are relatively unvaried referring to the much smaller numbers of immune cells.

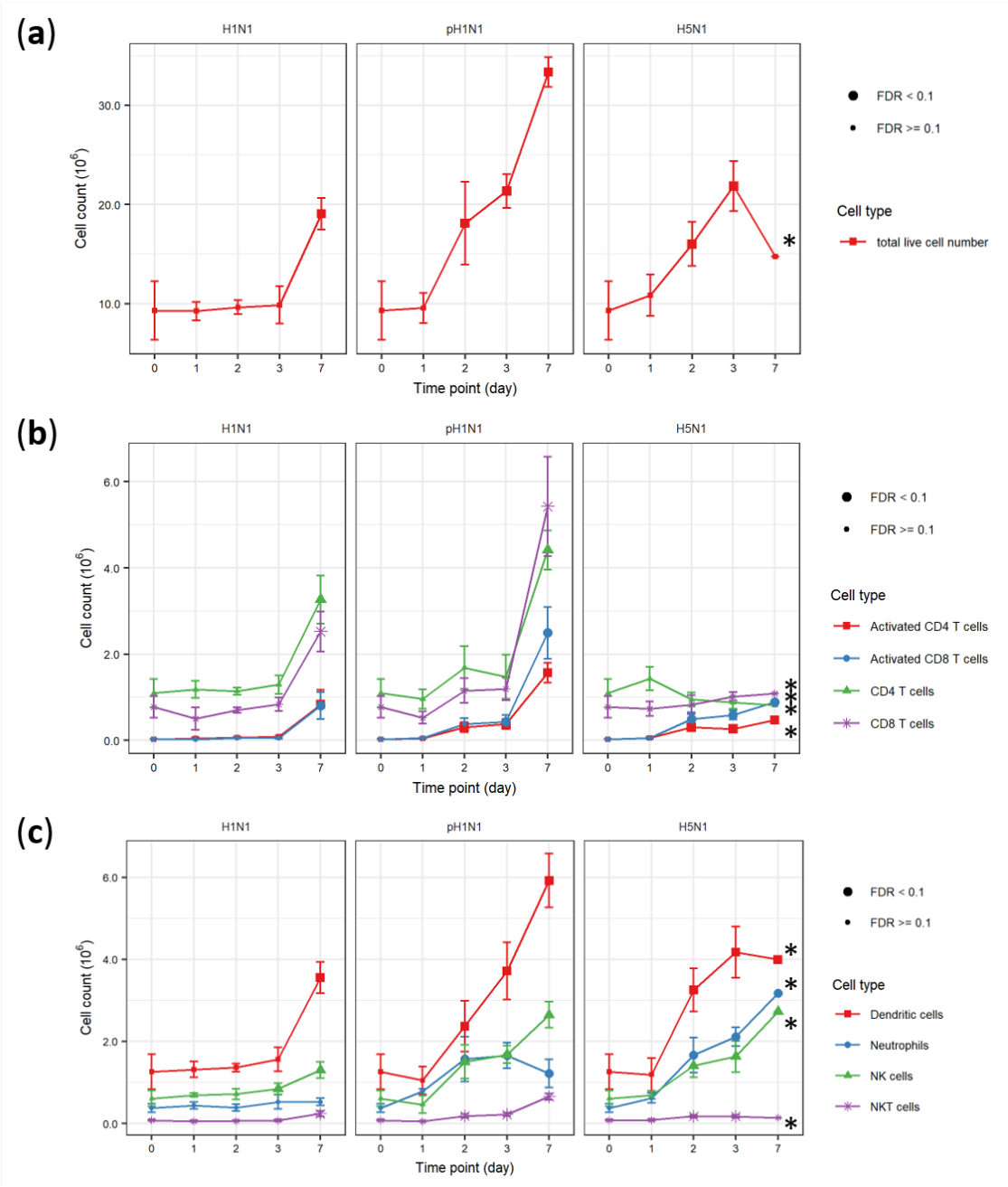


Figure 19: Cell counts of (a) total live cells, (b) T cell subsets, (c) neutrophils, DCs, NK cells and NKT cells in mouse lungs after infection by either H1N1, pH1N1 or H5N1 virus. Day 0 data are from uninfected, control animals. *Animals infected by H5N1 died before Day 7.

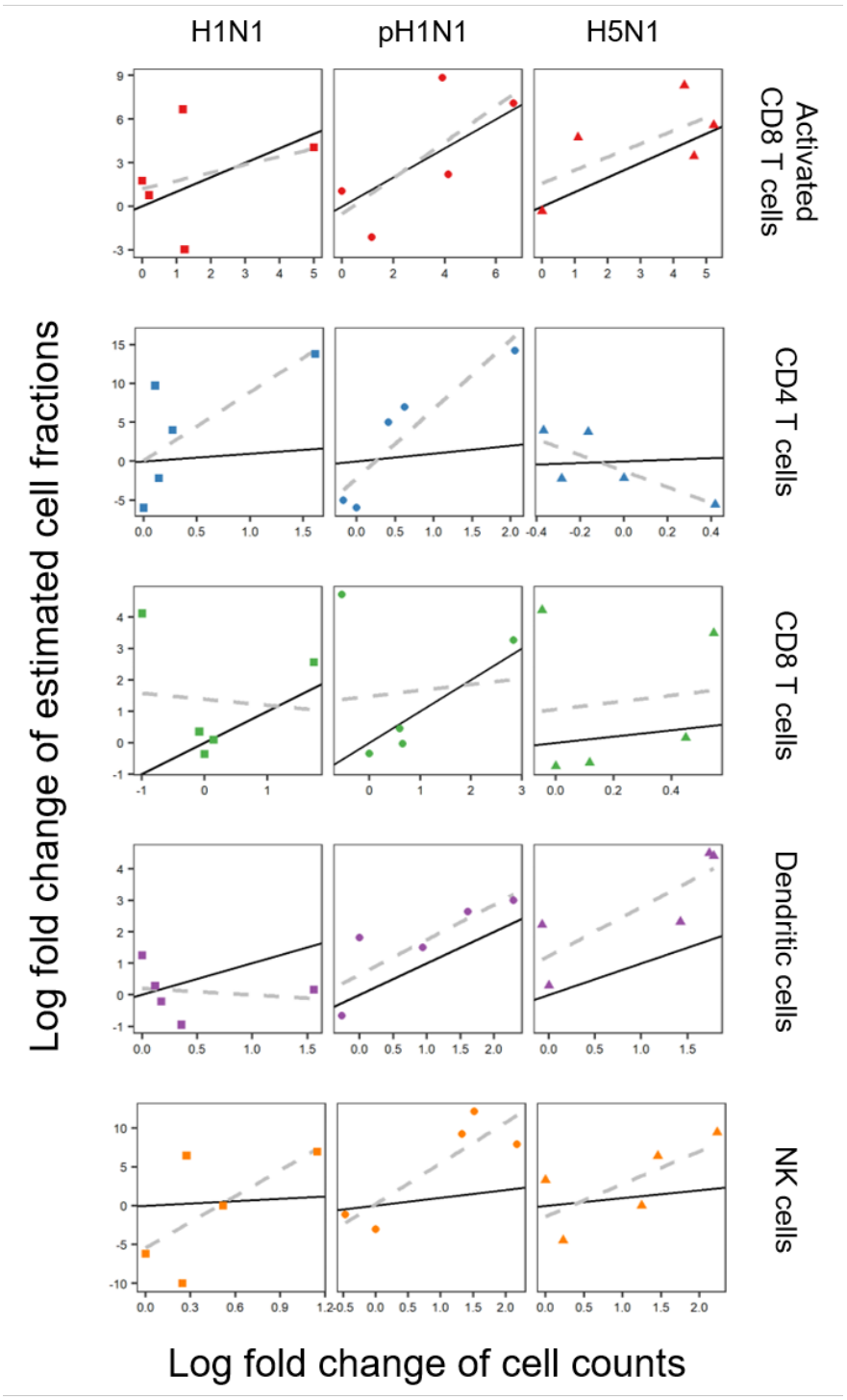


Figure 20: Log fold change of estimated cell fractions by CIBERSORT in comparison with log fold change of cell counts measured by FACS for different cell types. The black line is y = x while the grey dashed line is regression.

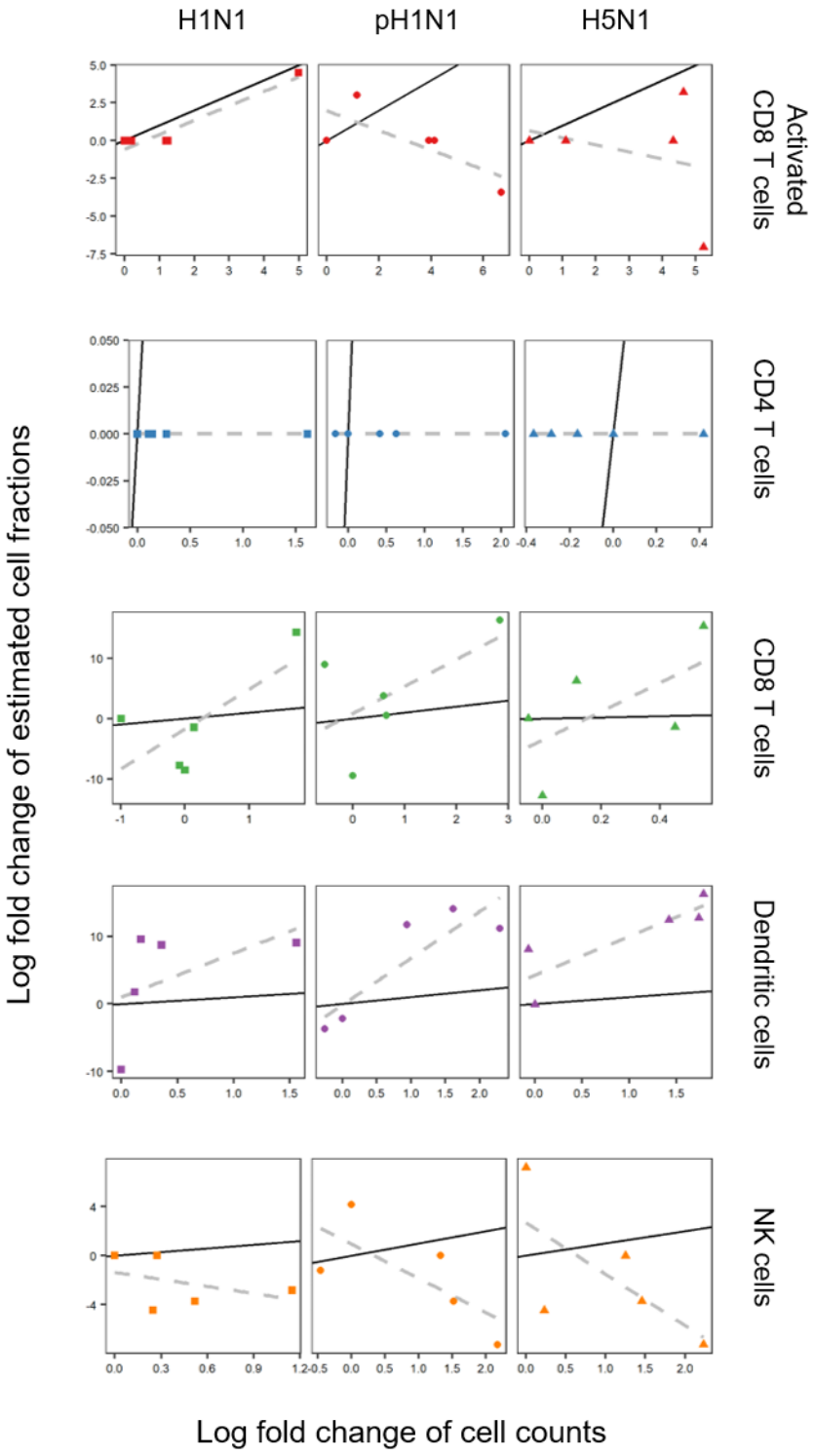


Figure 21: Log fold change of estimated cell fractions by MLLSR in comparison with log fold change of cell counts measured by FACS for different cell types. The black line is y = x while the grey dashed line is regression.

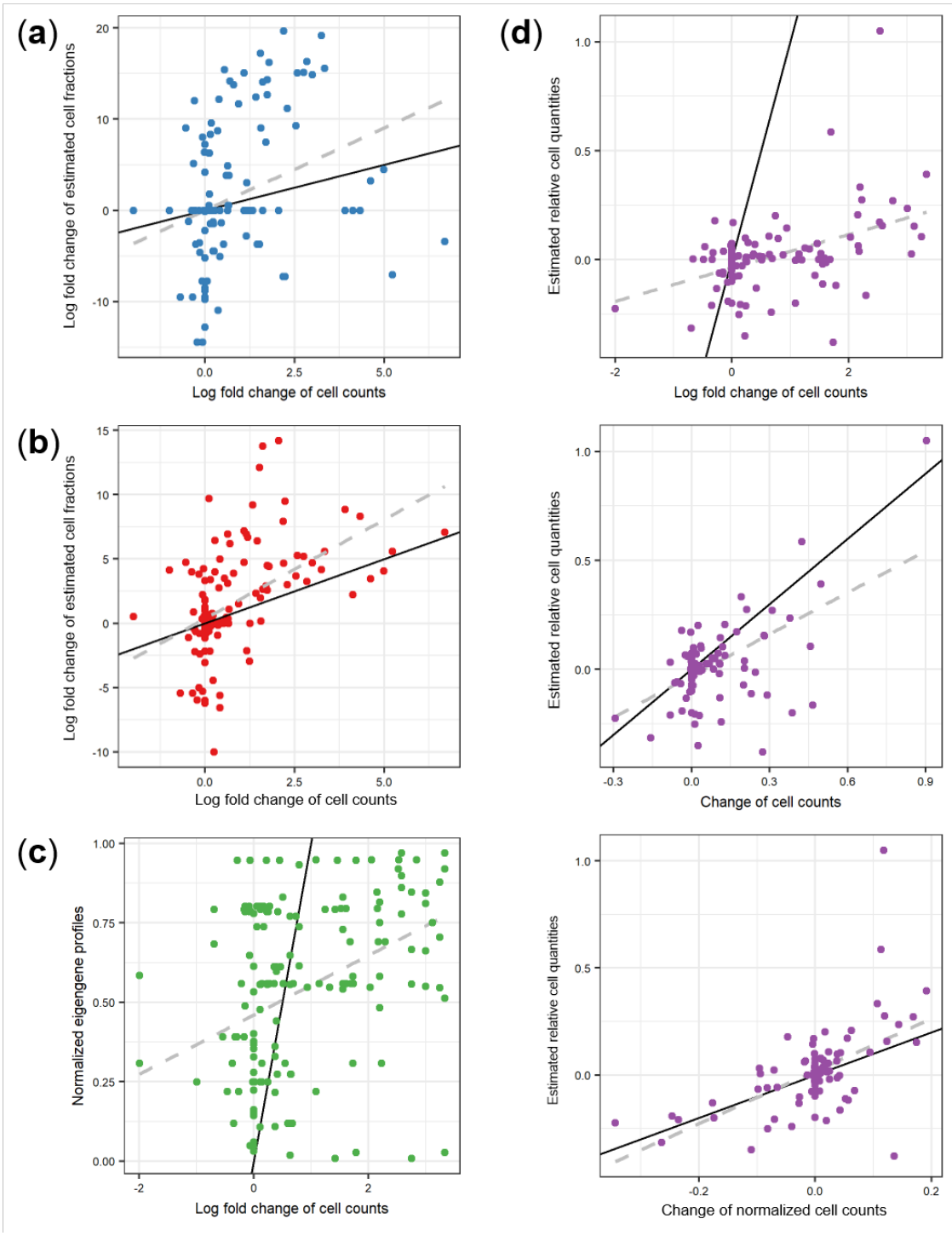


Figure 22: Estimated cell quantities by (a) MLLSR, (b) CIBERSORT, (c) CTen, and (d) DCQ in comparison with normalized cell counts measured by FACS. The black line is y = x while the grey dashed line is regression.

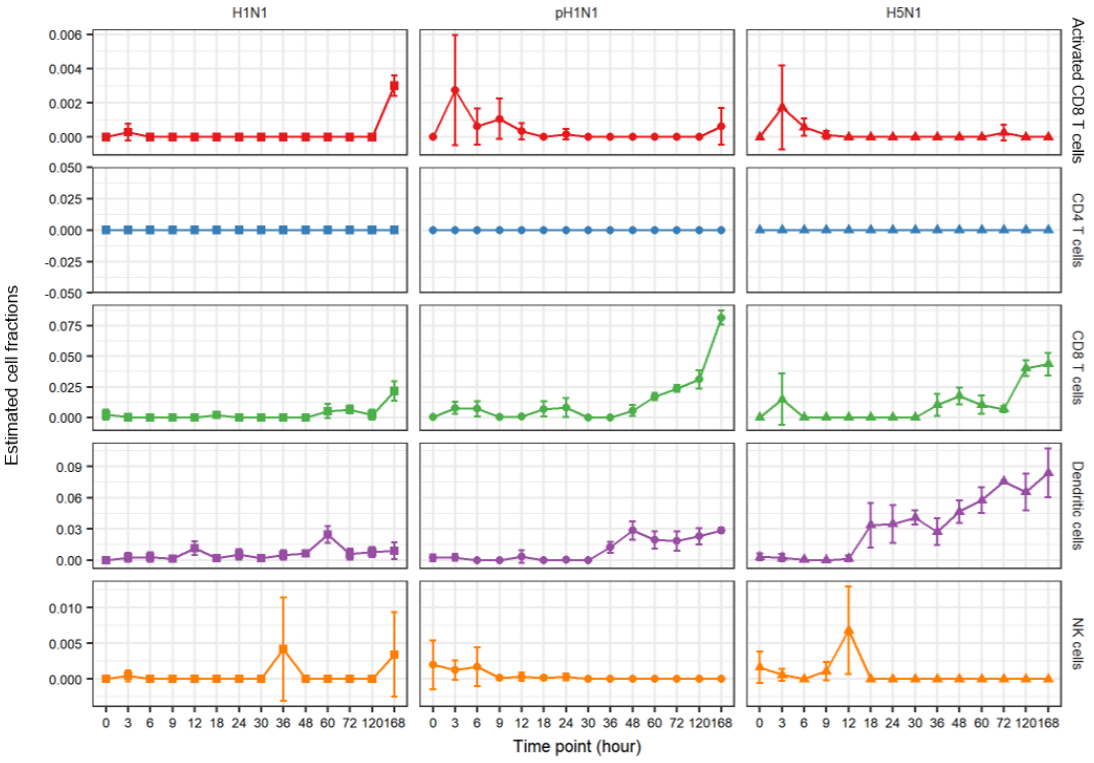


Figure 23: Estimated fractions of diverse cell types across time using MLLSR.

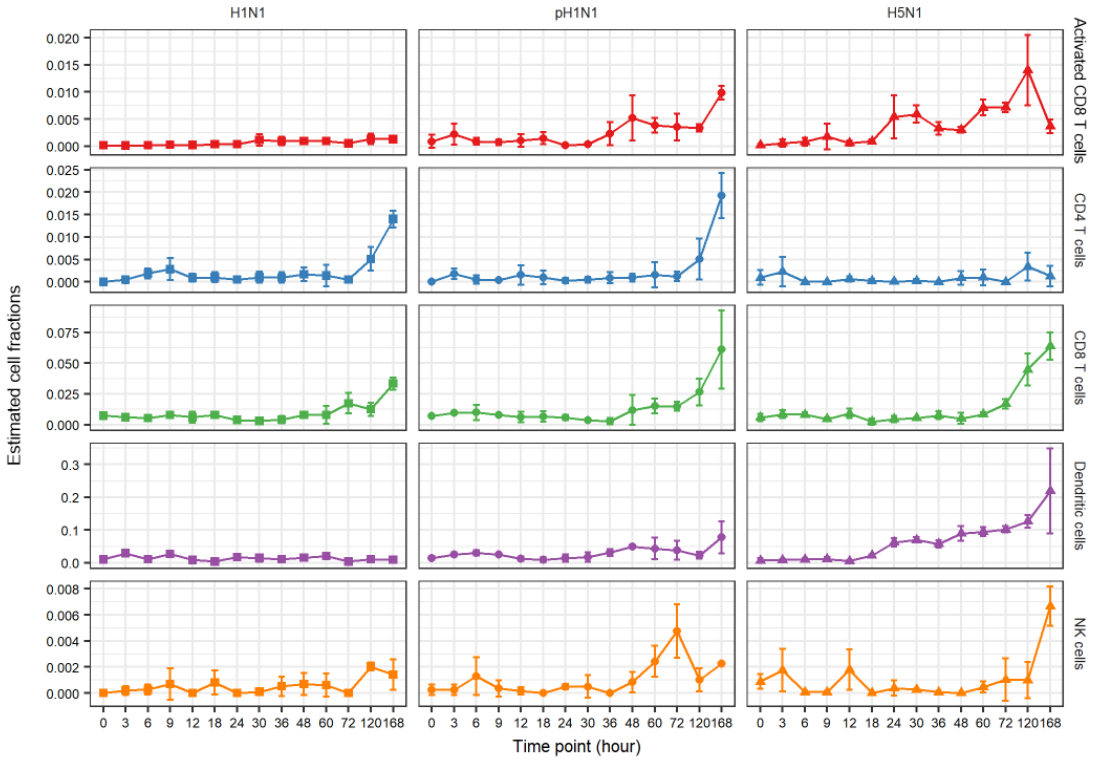


Figure 24: Estimated fractions of diverse cell types across time using CIBERSORT.

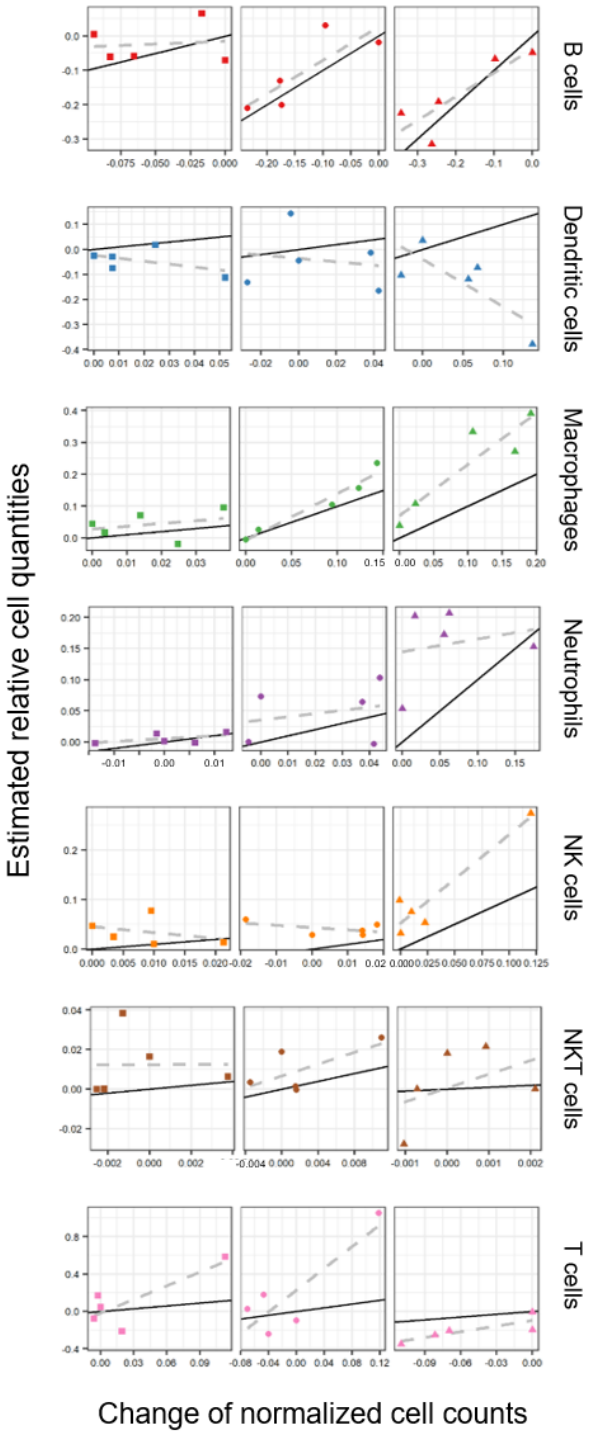


Figure 25: Estimated relative cell quantities by DCQ in comparison with the change in normalized cell counts for different cell types. The black line is y = x while the grey dashed line is regression.

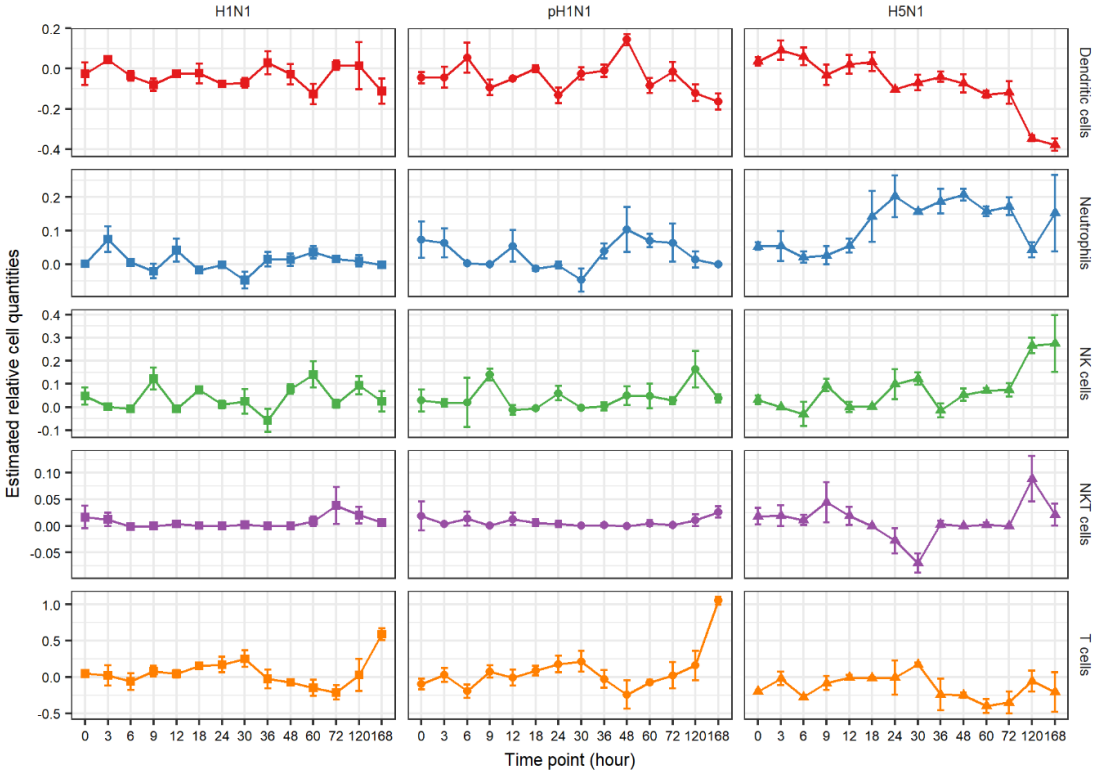


Figure 26: Estimated relative quantities of diverse cell types across time using DCQ.

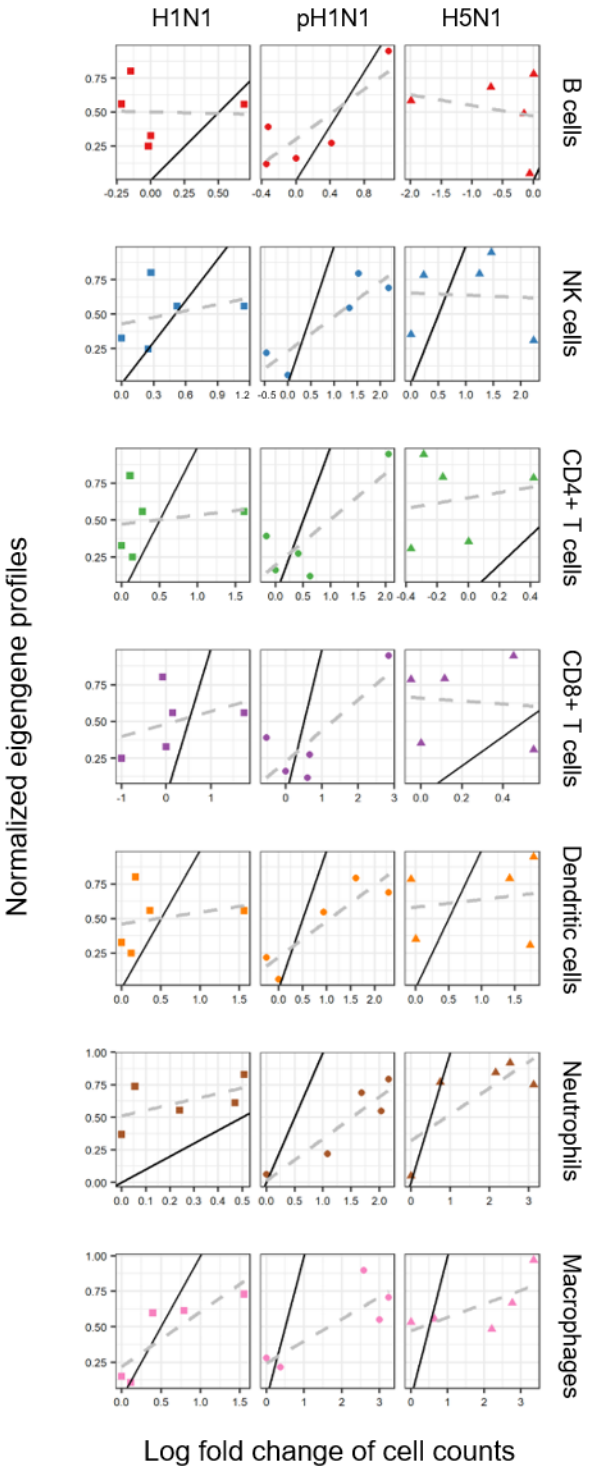


Figure 27: Normalized eigengene profiles by CTen WGCNA in comparison with log fold change of cell counts measured by FACS for different cell types. The black line is y = x while the grey dashed line is regression.

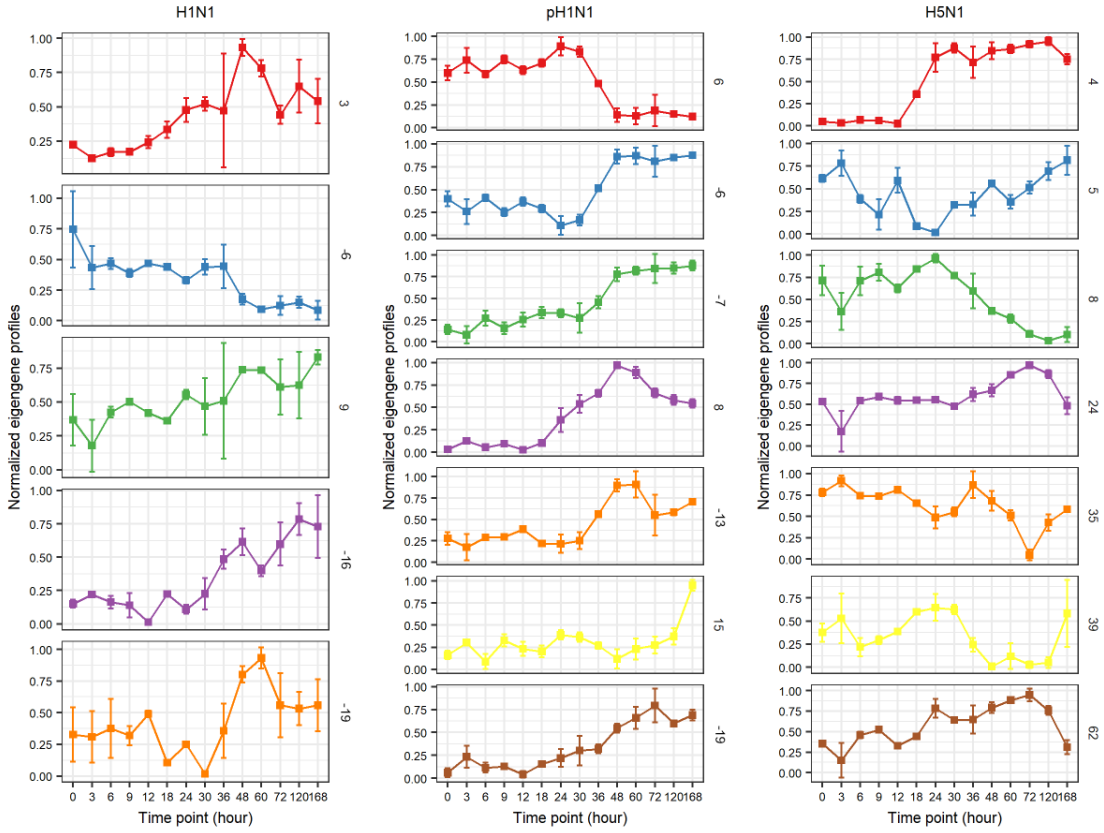


Figure 28: Normalized eigengene profiles of modules that are enriched in immune cells. Negative submodules are denoted by an extra minus sign.

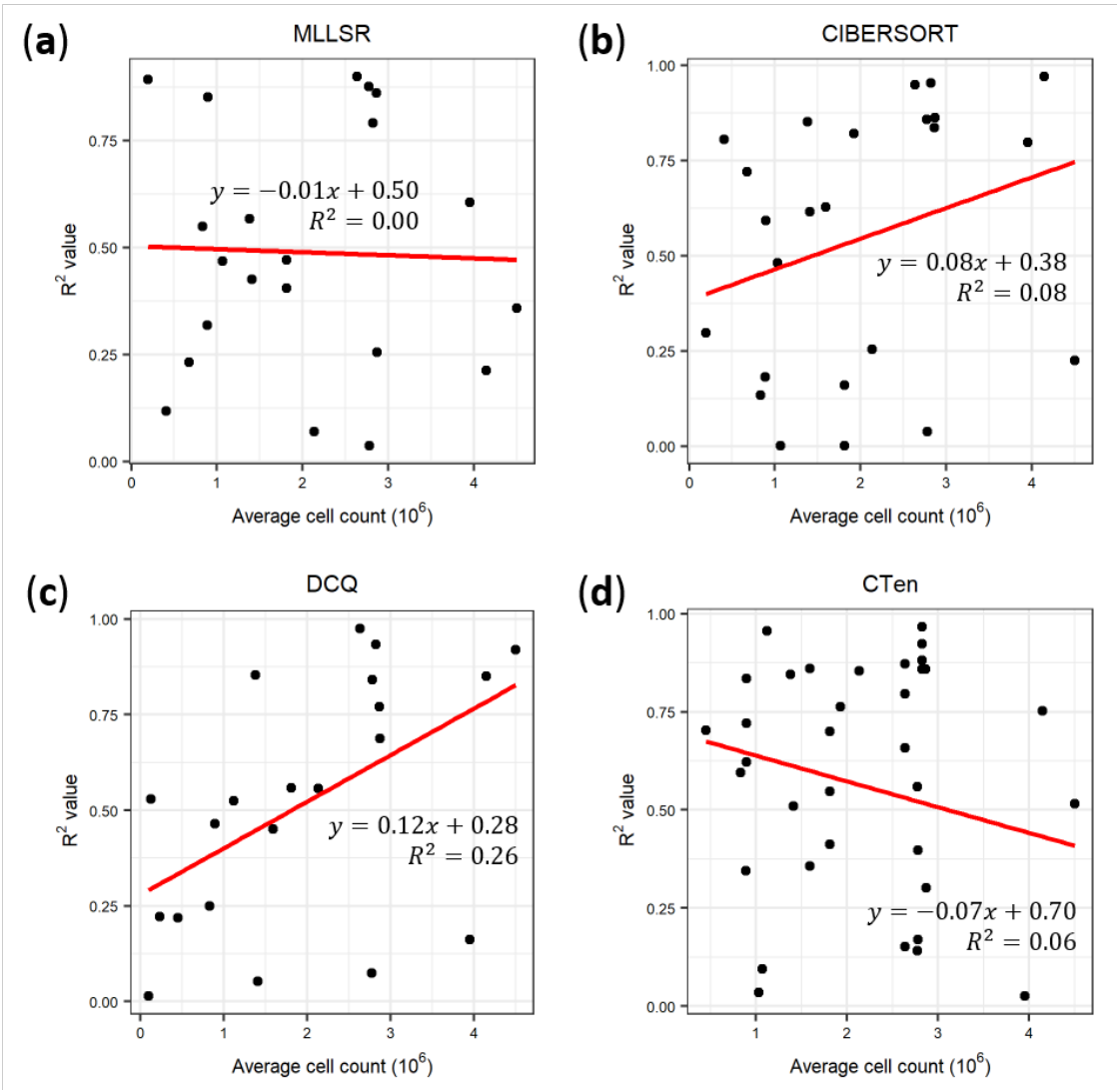


Figure 29: R^2 values obtained by comparing predicted cell quantities with measured cell counts are plotted against the average cell counts per cell type per sample cohort for each algorithm: (a) MLLSR, (b) CIBERSORT, (c) DCQ (p-value = 0.02), and (d) CTen. The red lines are regression lines with equations notated.

Table 2: Antibodies used in flow cytometry for each cell type. The signs "+" and "-" indicate positive and negative, respectively, and "H" and "L" represent "High" and "Low".

Cell type	CD45	c-kit	FceRIa	DX5	CD11b	CCR3	Gr-1	NK1.1	CD3	B220	CD4	CD8	F4/80	CD11c
Basophils	+	-	+	+										
Eosinophils	+				+	+	L							
Neutrophils	+				+		H							
B cells	+							-	-	+				
NK cells	+							+	-					
NK T cells	+							+	+					
CD4+ T cells	+								+		+	-		
CD8+ T cells	+								+		-	+		
Macrophages	+				+ or -								+	+ or -
Dendritic cells	+				+ or -								-	+

Appendix B

Supplementary Materials for Drug Target Inference

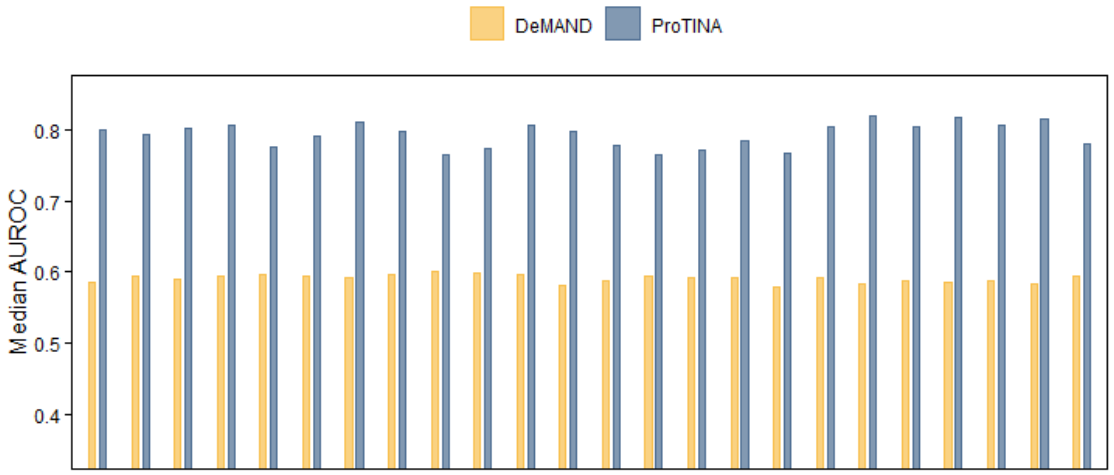


Figure 30: Comparing the accuracy of ProTINA and DeMAND. Target predictions for DP14 were performed by ProTINA and DeMAND using different combinations of PPIs and PGIs. Median AUROC values of both algorithms for the same network setup were paired in this figure.

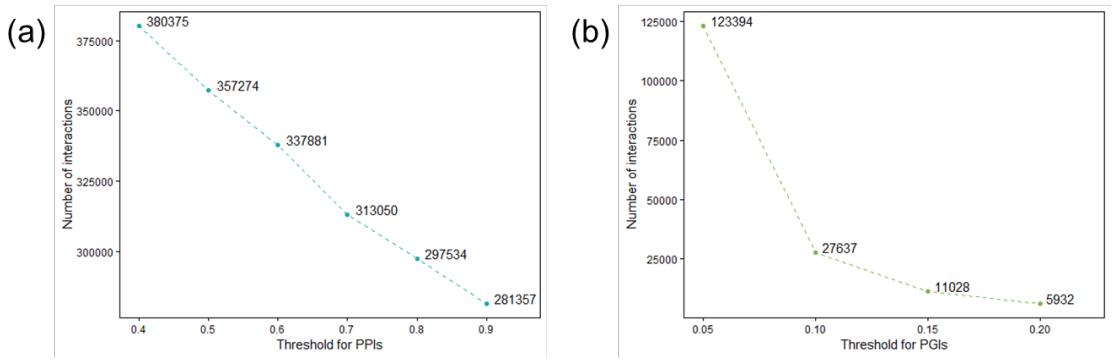


Figure 31: The number of interactions under different thresholds for (a) PPIs and (b) PGIs.

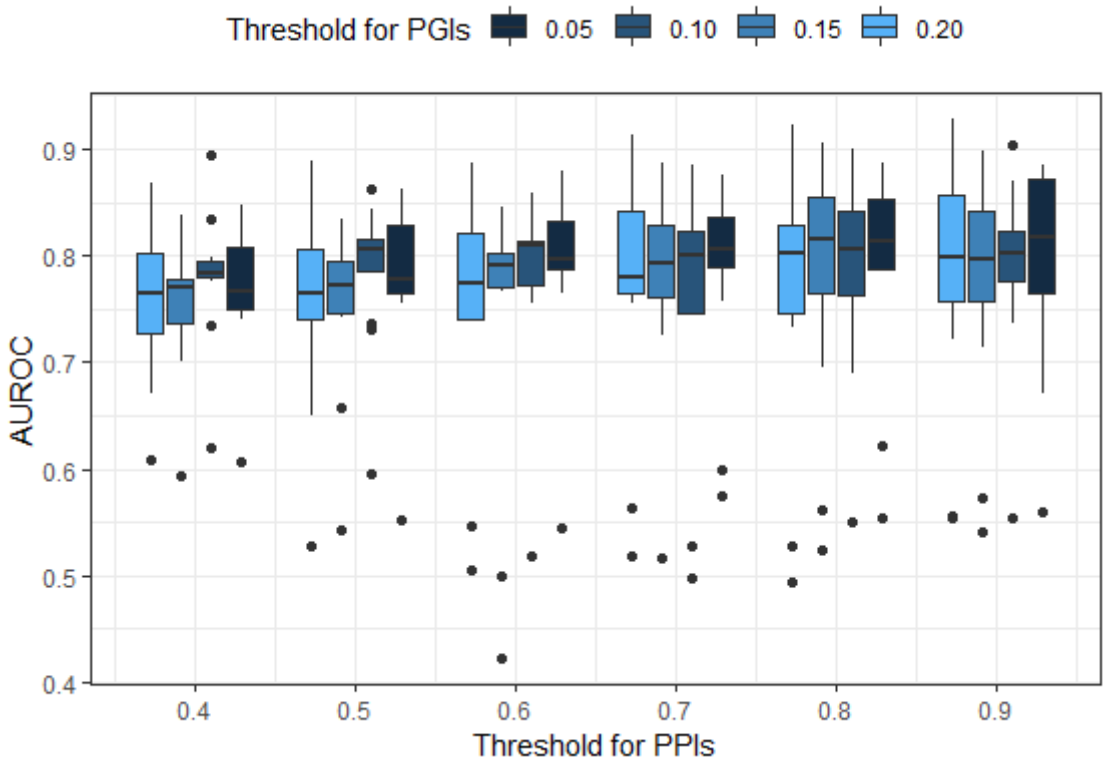


Figure 32: Performance of ProTINA on DP14 using PPIs and PGIs of different sizes.

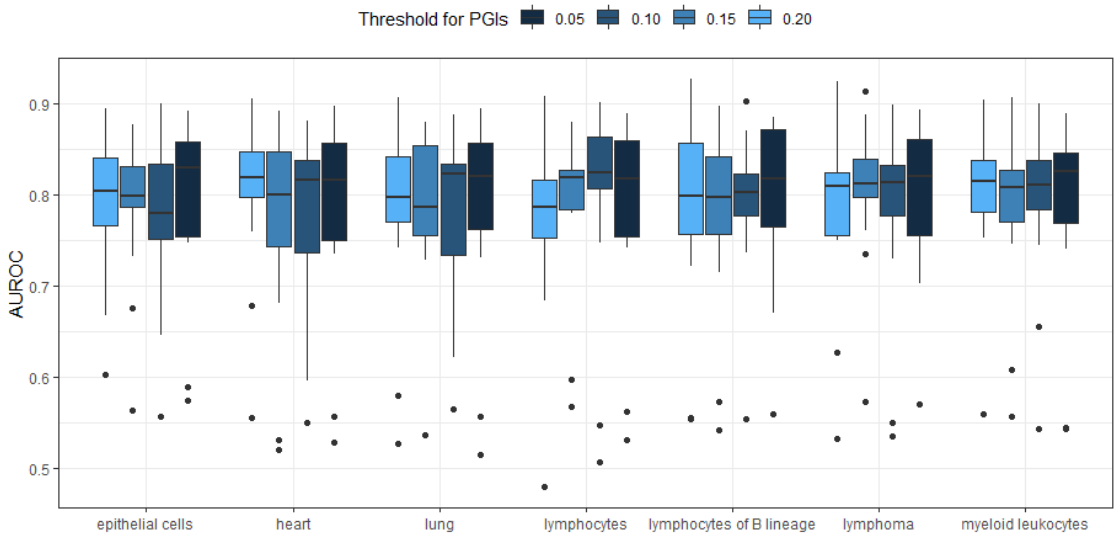


Figure 33: Performance of ProTINA on DP14 using PGIs from seven different cell/tissue types.

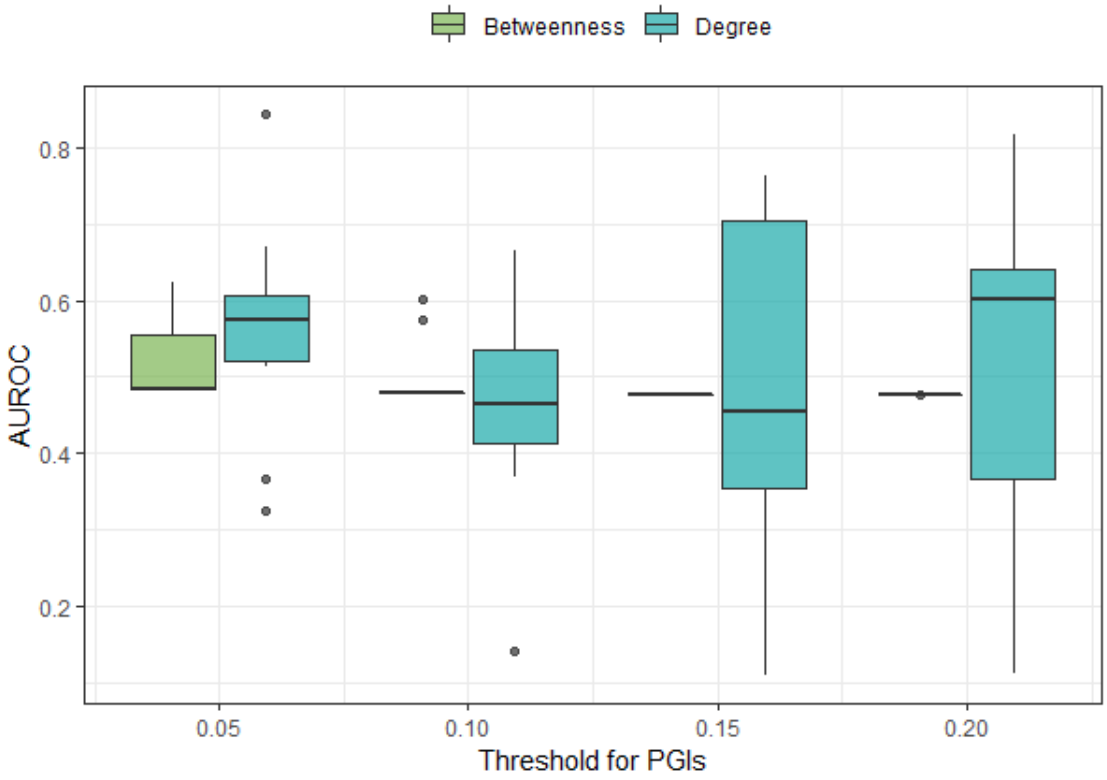


Figure 34: Predicting drug targets by degree or betweenness values of PGIs for 'lymphocytes of B lineage'.

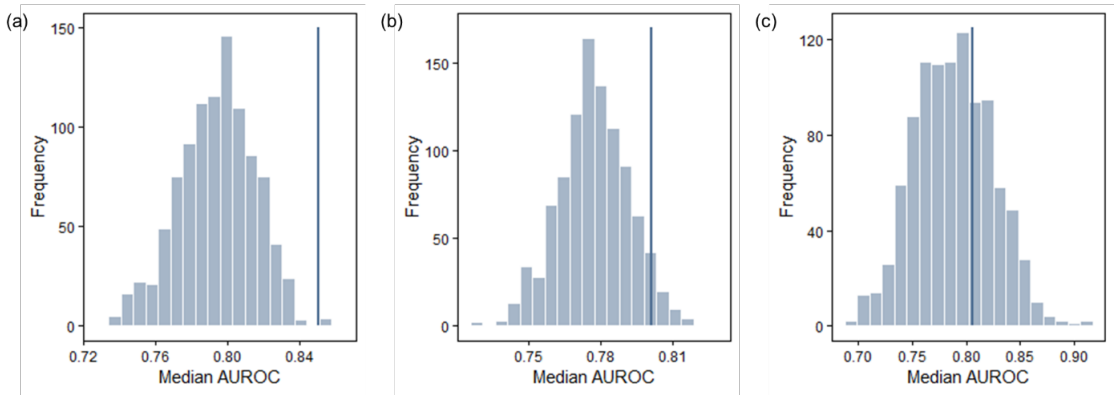


Figure 35: Median AUROCs of permutation tests on TREAP. 1000 permutation tests were performed on TREAP by randomizing the adjusted p-values for (a) DP14, (b) HepG2 and (c) MP. The blue vertical line in each panel refers to the median AUROC obtained by nonrandomized adjusted p-values.

Table 3: Correlation coefficients between degree or betweenness values and ProTINA scores for each drug

Chemical	PPI Betweenness	PPI Degree	PPI+PGI Betweenness	PPI+PGI Degree
Methotrexate	0.087742531	0.163785854	0.08539042	0.162703305
Blebbistatin	0.104272731	0.208298699	0.101154834	0.206385715
Camptothecin	0.106773529	0.224629718	0.102677995	0.22257972
Doxorubicin hydrochloride	0.112366047	0.209255068	0.114547587	0.208590487
Cycloheximide	0.106752819	0.214713019	0.103710782	0.212780846
Aclacinomycin A	0.024457165	0.097451912	0.023567461	0.096106315
Etoposide	0.08955877	0.178118355	0.085989173	0.176166631
Geldanamycin	0.06588576	0.169664995	0.068120054	0.168338047
Mitomycin C	0.094832663	0.219806504	0.092439142	0.217374173
Rapamycin	0.085112721	0.210553157	0.079170771	0.207861041
Trichostatin A	0.114241504	0.220097701	0.110610214	0.218378627
Vincristine	0.065358257	0.150465205	0.061458669	0.148480301
Monastrol	0.106149223	0.190908313	0.103970923	0.18895624
H-7, Dihydrochloride	0.099692622	0.228612182	0.099731216	0.227556513

Bibliography

[1] M. Wang, S. Fukuyama, Y. Kawaoka, and J. E. Shoemaker, "Predicting Host Immune Cell Dynamics Using Tissue Gene Expression," *IFAC-PapersOnLine*, vol. 51, no. 19, pp. 5–6, 2018.

[2] M. Wang, S. Fukuyama, Y. Kawaoka, and J. E. Shoemaker, "Predicting Host Immune Cell Dynamics and Key Disease-Associated Genes Using Tissue Transcriptional Profiles," *Processes*, vol. 7, no. 5, p. 301, 2019.

[3] R. Sato, C. Makino-Okamura, Q. Lin, M. Wang, J. E. Shoemaker, T. Kurosaki, and H. Fukuyama, "Repurposing the Psoriasis Drug Oxarol to an Ointment Adjuvant for the Influenza Vaccine," *International Immunology*, in press.

[4] M. Wang, H. Noh, E. Mochan, and J. E. Shoemaker, "Network Insights into Improving Drug Target Inference Algorithms," 2020.

[5] R. Macarron, M. N. Banks, D. Bojanic, D. J. Burns, D. A. Cirovic, T. Garyantes, D. V. S. Green, R. P. Hertzberg, W. P. Janzen, J. W. Paslay, U. Schopfer, and G. S. Sittampalam, "Impact of high-throughput screening in biomedical research," *Nature Reviews Drug Discovery*, vol. 10, no. 3, pp. 188–195, 2011.

[6] J. W. Scannell, A. Blanckley, H. Boldon, and B. Warrington, "Diagnosing the decline in pharmaceutical R&D efficiency," *Nat Rev Drug Discov*, vol. 11, no. 3, pp. 191–200, 2012.

[7] U.S. Food Drug Administration, "The Drug Development Process." https://www.fda.gov/patients/learn-about-drug-and-device-approvals/drug-development-process, 2018.

[8] U.S. Food Drug Administration, "Step 1: Discovery and Development." https://www.fda.gov/patients/drug-development-process/step-1-discovery-and-development, 2018.

[9] U.S. Food Drug Administration, "Step 2: Preclinical Research." https://www.fda.gov/patients/drug-development-process/step-2-preclinical-research, 2018.

[10] U.S. Food Drug Administration, "Step 3: Clinical Research." https://www.fda.gov/patients/drug-development-process/step-3-clinical-research, 2018.

[11] U.S. Food Drug Administration, "Step 4: FDA Drug Review." https://www.fda.gov/patients/drug-development-process/step-4-fda-drug-review, 2018.

[12] R. K. Harrison, "Phase II and phase III failures: 2013-2015," *Nat Rev Drug Discov*, vol. 15, no. 12, pp. 817–818, 2016.

[13] J. Arrowsmith, "Phase II failures: 2008–2010," *Nature Reviews Drug Discovery*, vol. 10, no. 5, pp. 328–329, 2011.

[14] J. Arrowsmith and P. Miller, "Phase II and Phase III attrition rates 2011–2012," *Nature Reviews Drug Discovery*, vol. 12, no. 8, pp. 569–569, 2013.

[15] U.S. Food and Drug Administration, Center for Drug Evaluation and Research, "CY 2018 CDER New Molecular Entity (NME) Drug & Original BLA Calendar Year Approvals," report, 2019.

[16] U.S. Food and Drug Administration, Center for Drug Evaluation and Research, "CY 2017 CDER New Molecular Entity (NME) Drug & Original BLA Calendar Year Approvals," report, 2019.

[17] U.S. Food and Drug Administration, Center for Drug Evaluation and Research, "CY 2016 CDER New Molecular Entity (NME) Drug & Original BLA Calendar Year Approvals," report, 2019.

[18] U.S. Food and Drug Administration, Center for Drug Evaluation and Research, "CY 2015 CDER New Molecular Entity (NME) Drug & Original BLA Calendar Year Approvals," report, 2019.

[19] J. P. Hughes, S. Rees, S. B. Kalindjian, and K. L. Philpott, "Principles of early drug discovery," *Br J Pharmacol*, vol. 162, no. 6, pp. 1239–49, 2011.

[20] P. Dorr, M. Westby, S. Dobbs, P. Griffin, B. Irvine, M. Macartney, J. Mori, G. Rickett, C. Smith-Burchnell, C. Napier, R. Webster, D. Armour, D. Price, B. Stammen, A. Wood, and M. Perros, "Maraviroc (UK-427,857), a potent, orally bioavailable, and selective small-molecule inhibitor of chemokine receptor CCR5 with broad-spectrum

anti-human immunodeficiency virus type 1 activity," *Antimicrob Agents Chemother*, vol. 49, no. 11, pp. 4721–32, 2005.

[21] D. Cook, D. Brown, R. Alexander, R. March, P. Morgan, G. Satterthwaite, and M. N. Pangalos, "Lessons learned from the fate of AstraZeneca's drug pipeline: a five-dimensional framework," *Nat Rev Drug Discov*, vol. 13, no. 6, pp. 419–31, 2014.

[22] K. J. Duffy, A. N. Shaw, E. Delorme, S. B. Dillon, C. Erickson-Miller, L. Giampa, Y. Huang, R. M. Keenan, P. Lamb, N. Liu, S. G. Miller, A. T. Price, J. Rosen, H. Smith, K. J. Wiggall, L. Zhang, and J. I. Luengo, "Identification of a pharmacophore for thrombopoietic activity of small, non-peptidyl molecules. 1. Discovery and optimization of salicylaldehyde thiosemicarbazone thrombopoietin mimics," *J Med Chem*, vol. 45, no. 17, pp. 3573–5, 2002.

[23] M. Gao, R. E. Nettles, M. Belema, L. B. Snyder, V. N. Nguyen, R. A. Fridell, M. H. Serrano-Wu, D. R. Langley, J. H. Sun, n. O'Boyle, D. R., J. A. Lemm, C. Wang, J. O. Knipe, C. Chien, R. J. Colonno, D. M. Grasela, N. A. Meanwell, and L. G. Hamann, "Chemical genetics strategy identifies an HCV NS5A inhibitor with a potent clinical effect," *Nature*, vol. 465, no. 7294, pp. 96–100, 2010.

[24] J. Seok, H. S. Warren, A. G. Cuenca, M. N. Mindrinos, H. V. Baker, W. Xu, D. R. Richards, G. P. McDonald-Smith, H. Gao, L. Hennessy, C. C. Finnerty, C. M. Lopez, S. Honari, E. E. Moore, J. P. Minei, J. Cuschieri, P. E. Bankey, J. L. Johnson, J. Sperry, A. B. Nathens, T. R. Billiar, M. A. West, M. G. Jeschke, M. B. Klein, R. L. Gamelli, N. S. Gibran, B. H. Brownstein, C. Miller-Graziano, S. E. Calvano, P. H. Mason, J. P. Cobb, L. G. Rahme, S. F. Lowry, R. V. Maier, L. L. Moldawer, D. N. Herndon, R. W. Davis, W. Xiao, R. G. Tompkins, Inflammation, and L. S. C. R. P. Host Response to Injury, "Genomic responses in mouse models poorly mimic human inflammatory diseases," *Proc Natl Acad Sci U S A*, vol. 110, no. 9, pp. 3507–12, 2013.

[25] M. de Jong and T. Maina, "Of mice and humans: are they the same?–Implications in cancer translational research," *J Nucl Med*, vol. 51, no. 4, pp. 501–4, 2010.

[26] A. Wendler and M. Wehling, "The translatability of animal models for clinical development: biomarkers and disease models," *Curr Opin Pharmacol*, vol. 10, no. 5, pp. 601–6, 2010.

[27] M. Jucker, "The benefits and limitations of animal models for translational research in neurodegenerative diseases," *Nat Med*, vol. 16, no. 11, pp. 1210–4, 2010.

[28] R. Lowe, N. Shirley, M. Bleackley, S. Dolan, and T. Shafee, "Transcriptomics technologies," *PLoS Comput Biol*, vol. 13, no. 5, p. e1005457, 2017.

[29] J. E. Shoemaker, S. Fukuyama, A. J. Eisfeld, D. Zhao, E. Kawakami, S. Sakabe, T. Maemura, T. Gorai, H. Katsura, Y. Muramoto, S. Watanabe, T. Watanabe, K. Fuji, Y. Matsuoka, H. Kitano, and Y. Kawaoka, "An ultrasensitive mechanism regulates influenza virus-induced inflammation," *PLoS Pathog*, vol. 11, no. 6, p. e1004856, 2015.

[30] J. H. Woo, Y. Shimoni, W. S. Yang, P. Subramaniam, A. Iyer, P. Nicoletti, M. Rodriguez Martinez, G. Lopez, M. Mattioli, R. Realubit, C. Karan, B. R. Stockwell, M. Bansal, and A. Califano, "Elucidating compound mechanism of action by network perturbation analysis," *Cell*, vol. 162, no. 2, pp. 441–451, 2015.

[31] H. Noh, J. E. Shoemaker, and R. Gunawan, "Network perturbation analysis of gene transcriptional profiles reveals protein targets and mechanism of action of drugs and influenza A viral infection," *Nucleic Acids Res*, 2018.

[32] M. Santolini and A. L. Barabasi, "Predicting perturbation patterns from the topology of biological networks," *Proc Natl Acad Sci U S A*, vol. 115, no. 27, pp. E6375–E6383, 2018.

[33] S. Jain, J. Arrais, N. J. Venkatachari, V. Ayyavoo, and Z. Bar-Joseph, "Reconstructing the temporal progression of HIV-1 immune response pathways," *Bioinformatics*, vol. 32, no. 12, pp. i253–i261, 2016.

[34] B. B. Cummings, J. L. Marshall, T. Tukiainen, M. Lek, S. Donkervoort, A. R. Foley, V. Bolduc, L. B. Waddell, S. A. Sandaradura, G. L. O'Grady, E. Estrella, H. M. Reddy, F. Zhao, B. Weisburd, K. J. Karczewski, A. H. O'Donnell-Luria, D. Birnbaum, A. Sarkozy, Y. Hu, H. Gonorazky, K. Claeys, H. Joshi, A. Bournazos, E. C. Oates, R. Ghaoui, M. R. Davis, N. G. Laing, A. Topf, C. Genotype-Tissue Expression, P. B. Kang, A. H. Beggs, K. N. North, V. Straub, J. J. Dowling, F. Muntoni, N. F. Clarke, S. T. Cooper, C. G. Bonnemann, and D. G. MacArthur, "Improving genetic diagnosis in Mendelian disease with transcriptome sequencing," *Sci Transl Med*, vol. 9, no. 386, 2017.

[35] Z. Isik, C. Baldow, C. V. Cannistraci, and M. Schroeder, "Drug target prioritization by perturbed gene expression and network information," *Sci Rep*, vol. 5, p. 17417, 2015.

[36] H. Nair, W. A. Brooks, M. Katz, A. Roca, J. A. Berkley, S. A. Madhi, J. M. Simmerman, A. Gordon, M. Sato, S. Howie, A. Krishnan, M. Ope, K. A. Lindblade, P. Carosone-Link, M. Lucero, W. Ochieng, L. Kamimoto, E. Dueger, N. Bhat, S. Vong, E. Theodoratou, M. Chittaganpitch, O. Chimah, A. Balmaseda, P. Buchy, E. Harris, V. Evans, M. Katayose, B. Gaur, C. O'Callaghan-Gordo, D. Goswami, W. Arvelo, M. Venter, T. Briese, R. Tokarz, M.-A. Widdowson, A. W. Mounts, R. F. Breiman, D. R. Feikin, K. P. Klugman, S. J. Olsen, B. D. Gessner, P. F. Wright, I. Rudan, S. Broor, E. A. F. Simões, and H. Campbell, "Global burden of respiratory infections due to seasonal influenza in young children: a systematic review and meta-analysis," *The Lancet*, vol. 378, no. 9807, pp. 1917–1930, 2011.

[37] W. W. Thompson, E. Weintraub, P. Dhankhar, P.-Y. Cheng, L. Brammer, M. I. Meltzer, J. S. Bresee, and D. K. Shay, "Estimates of US influenza-associated deaths made using four different methods," *Influenza and Other Respiratory Viruses*, vol. 3, no. 1, pp. 37–49, 2009.

[38] Centers for Disease Control and Prevention, National Center for Immunization and Respiratory Diseases (NCIRD), "2017-2018 Estimated Influenza Illnesses, Medical visits, Hospitalizations, and Deaths and Estimated Influenza Illnesses, Medical visits, Hospitalizations, and Deaths Averted by Vaccination in the United States." https://www.cdc.gov/flu/about/burden-averted/2017-2018.htm, 2019.

[39] B. Li, V. Ruotti, R. M. Stewart, J. A. Thomson, and C. N. Dewey, "RNA-Seq gene expression estimation with read mapping uncertainty," *Bioinformatics*, vol. 26, no. 4, pp. 493–500, 2009.

[40] M. D. Robinson and A. Oshlack, "A scaling normalization method for differential expression analysis of RNA-seq data," *Genome Biology*, vol. 11, no. 3, p. R25, 2010.

[41] W. Huang da, B. T. Sherman, and R. A. Lempicki, "Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources," *Nat Protoc*, vol. 4, no. 1, pp. 44–57, 2009.

[42] W. Huang da, B. T. Sherman, and R. A. Lempicki, "Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists," *Nucleic Acids Res*, vol. 37, no. 1, pp. 1–13, 2009.

[43] M. S. Mackroth, A. Abel, C. Steeg, J. Schulze Zur Wiesch, and T. Jacobs, "Acute Malaria Induces PD1+CTLA4+ Effector T Cells with Cell-Extrinsic Suppressor Function," *PLoS Pathog*, vol. 12, no. 11, p. e1005909, 2016.

[44] D. Ostroumov, N. Fekete-Drimusz, M. Saborowski, F. Kuhnel, and N. Woller, "CD4 and CD8 T lymphocyte interplay in controlling tumor growth," *Cell Mol Life Sci*, vol. 75, no. 4, pp. 689–713, 2018.

[45] C. A. Dendrou, L. Fugger, and M. A. Friese, "Immunopathology of multiple sclerosis," *Nat Rev Immunol*, vol. 15, no. 9, pp. 545–58, 2015.

[46] L. Josset, J. A. Belser, M. J. Pantin-Jackwood, J. H. Chang, S. T. Chang, S. E. Belisle, T. M. Tumpey, and M. G. Katze, "Implication of inflammatory macrophages, nuclear receptors, and interferon regulatory factors in increased virulence of pandemic 2009 H1N1 influenza A virus after host adaptation," *J Virol*, vol. 86, no. 13, pp. 7192–206, 2012.

[47] R. Warrington, W. Watson, H. L. Kim, and F. R. Antonetti, "An introduction to immunology and immunopathology," *Allergy Asthma Clin Immunol*, vol. 7 Suppl 1, p. S1, 2011.

[48] H. I. Nakaya, J. Wrammert, E. K. Lee, L. Racioppi, S. Marie-Kunze, W. N. Haining, A. R. Means, S. P. Kasturi, N. Khan, G. M. Li, M. McCausland, V. Kanchan, K. E. Kokko, S. Li, R. Elbein, A. K. Mehta, A. Aderem, K. Subbarao, R. Ahmed, and B. Pulendran, "Systems biology of vaccination for seasonal influenza in humans," *Nat Immunol*, vol. 12, no. 8, pp. 786–95, 2011.

[49] G. Obermoser, S. Presnell, K. Domico, H. Xu, Y. Wang, E. Anguiano, L. Thompson-Snipes, R. Ranganathan, B. Zeitner, A. Bjork, D. Anderson, C. Speake, E. Ruchaud, J. Skinner, L. Alsina, M. Sharma, H. Dutartre, A. Cepika, E. Israelsson, P. Nguyen, Q. A. Nguyen, A. C. Harrod, S. M. Zurawski, V. Pascual, H. Ueno, G. T. Nepom, C. Quinn, D. Blankenship, K. Palucka, J. Banchereau, and D. Chaussabel, "Systems scale interactive exploration reveals quantitative and qualitative differences in response to influenza and pneumococcal vaccines," *Immunity*, vol. 38, no. 4, pp. 831–44, 2013.

[50] S. M. Kaech and R. Ahmed, "Memory CD8+ T cell differentiation: initial antigen encounter triggers a developmental program in naive cells," *Nat Immunol*, vol. 2, no. 5, pp. 415–22, 2001.

[51] A. D. Luster, R. Alon, and U. H. von Andrian, "Immune cell migration in inflammation: present and future therapeutic targets," *Nat Immunol*, vol. 6, no. 12, pp. 1182–90, 2005.

[52] P. A. Lang, K. S. Lang, H. C. Xu, M. Grusdat, I. A. Parish, M. Recher, A. R. Elford, S. Dhanji, N. Shaabani, C. W. Tran, D. Dissanayake, R. Rahbar, M. Ghazarian, A. Brustle, J. Fine, P. Chen, C. T. Weaver, C. Klose, A. Diefenbach, D. Haussinger, J. R. Carlyle, S. M. Kaech, T. W. Mak, and P. S. Ohashi, "Natural killer cell activation enhances immune pathology and promotes chronic infection by limiting CD8+ T-cell immunity," *Proc Natl Acad Sci U S A*, vol. 109, no. 4, pp. 1210–5, 2012.

[53] V. C. Lam and L. L. Lanier, "NK cells in host responses to viral infections," *Curr Opin Immunol*, vol. 44, pp. 43–51, 2017.

[54] J. Morrison, L. Josset, N. Tchitchek, J. Chang, J. A. Belser, D. E. Swayne, M. J. Pantin-Jackwood, T. M. Tumpey, and M. G. Katze, "H7N9 and other pathogenic avian influenza viruses elicit a three-pronged transcriptomic signature that is reminiscent of 1918 influenza virus and is associated with lethal outcome in mice," *J Virol*, vol. 88, no. 18, pp. 10556–68, 2014.

[55] J. S. Peiris, C. Y. Cheung, C. Y. Leung, and J. M. Nicholls, "Innate immune responses to influenza A H5N1: friend or foe?," *Trends Immunol*, vol. 30, no. 12, pp. 574–84, 2009.

[56] C. Cilloniz, K. Shinya, X. Peng, M. J. Korth, S. C. Proll, L. D. Aicher, V. S. Carter, J. H. Chang, D. Kobasa, F. Feldmann, J. E. Strong, H. Feldmann, Y. Kawaoka, and M. G. Katze, "Lethal influenza virus infection in macaques is associated with early dysregulation of inflammatory related genes," *PLoS Pathog*, vol. 5, no. 10, p. e1000604, 2009.

[57] M. Brandes, F. Klauschen, S. Kuchen, and R. N. Germain, "A systems analysis identifies a feedforward inflammatory circuit leading to lethal influenza infection," *Cell*, vol. 154, no. 1, pp. 197–212, 2013.

[58] M. J. Carter, "A rationale for using steroids in the treatment of severe cases of H5N1 avian influenza," *J Med Microbiol*, vol. 56, no. Pt 7, pp. 875–83, 2007.

[59] K. Shinya, M. Ito, A. Makino, M. Tanaka, K. Miyake, A. J. Eisfeld, and Y. Kawaoka, "The TLR4-TRIF pathway protects against H5N1 influenza virus infection," *J Virol*, vol. 86, no. 1, pp. 19–24, 2012.

[60] A. Tanaka, S. Nakamura, M. Seki, K. Fukudome, N. Iwanaga, Y. Imamura, T. Miyazaki, K. Izumikawa, H. Kakeya, K. Yanagihara, and S. Kohno, "Toll-like receptor 4 agonistic antibody promotes innate immunity against severe pneumonia

induced by coinfection with influenza virus and Streptococcus pneumoniae," *Clin Vaccine Immunol*, vol. 20, no. 7, pp. 977–85, 2013.

[61] S. F. Ibrahim and G. van den Engh, "Flow cytometry and cell sorting," *Adv Biochem Eng Biotechnol*, vol. 106, pp. 19–39, 2007.

[62] S. Bhattacharya, P. Dunn, C. G. Thomas, B. Smith, H. Schaefer, J. Chen, Z. Hu, K. A. Zalocusky, R. D. Shankar, S. S. Shen-Orr, E. Thomson, J. Wiser, and A. J. Butte, "ImmPort, toward repurposing of open access immunological assay data for translational and clinical research," *Sci Data*, vol. 5, p. 180015, 2018.

[63] R. Edgar, M. Domrachev, and A. E. Lash, "Gene Expression Omnibus: NCBI gene expression and hybridization array data repository," *Nucleic Acids Res*, vol. 30, no. 1, pp. 207–10, 2002.

[64] J. E. Shoemaker, T. J. Lopes, S. Ghosh, Y. Matsuoka, Y. Kawaoka, and H. Kitano, "CTen: a web-based platform for identifying enriched cell types from heterogeneous microarray data," *BMC Genomics*, vol. 13, p. 460, 2012.

[65] M. Wang, S. R. Master, and L. A. Chodosh, "Computational expression deconvolution in a complex mammalian organ," *BMC Bioinformatics*, vol. 7, p. 328, 2006.

[66] Y. Zhong, Y. W. Wan, K. Pang, L. M. Chow, and Z. Liu, "Digital sorting of complex tissues for cell type-specific gene expression profiles," *BMC Bioinformatics*, vol. 14, p. 89, 2013.

[67] A. R. Abbas, K. Wolslegel, D. Seshasayee, Z. Modrusan, and H. F. Clark, "Deconvolution of blood microarray data identifies cellular activation patterns in systemic lupus erythematosus," *PLoS One*, vol. 4, no. 7, p. e6098, 2009.

[68] P. Lu, A. Nakorchevskiy, and E. M. Marcotte, "Expression deconvolution: a reinterpretation of DNA microarray data reveals dynamic changes in cell populations," *Proc Natl Acad Sci U S A*, vol. 100, no. 18, pp. 10370–5, 2003.

[69] Z. Altboum, Y. Steuerman, E. David, Z. Barnett-Itzhaki, L. Valadarsky, H. Keren-Shaul, T. Meningher, E. Mendelson, M. Mandelboim, I. Gat-Viks, and I. Amit, "Digital cell quantification identifies global immune cell dynamics during influenza infection," *Mol Syst Biol*, vol. 10, p. 720, 2014.

[70] A. M. Newman, C. L. Liu, M. R. Green, A. J. Gentles, W. Feng, Y. Xu, C. D. Hoang, M. Diehn, and A. A. Alizadeh, "Robust enumeration of cell subsets from tissue expression profiles," *Nat Methods*, vol. 12, no. 5, pp. 453–7, 2015.

[71] D. A. Liebner, K. Huang, and J. D. Parvin, "MMAD: microarray microdissection with analysis of differences is a computational tool for deconvoluting cell type-specific contributions from tissue samples," *Bioinformatics*, vol. 30, no. 5, pp. 682–9, 2014.

[72] M. E. Ritchie, B. Phipson, D. Wu, Y. Hu, C. W. Law, W. Shi, and G. K. Smyth, "limma powers differential expression analyses for RNA-sequencing and microarray studies," *Nucleic Acids Research*, vol. 43, no. 7, pp. e47–e47, 2015.

[73] L. Escoubet-Lozach, C. Benner, M. U. Kaikkonen, J. Lozach, S. Heinz, N. J. Spann, A. Crotti, J. Stender, S. Ghisletti, D. Reichart, C. S. Cheng, R. Luna, C. Ludka, R. Sasik, I. Garcia-Bassets, A. Hoffmann, S. Subramaniam, G. Hardiman, M. G. Rosenfeld, and C. K. Glass, "Mechanisms establishing TLR4-responsive activation states of inflammatory response genes," *PLoS Genet*, vol. 7, no. 12, p. e1002401, 2011.

[74] F. Pan, H. Yu, E. V. Dang, J. Barbi, X. Pan, J. F. Grosso, D. Jinasena, S. M. Sharma, E. M. McCadden, D. Getnet, C. G. Drake, J. O. Liu, M. C. Ostrowski, and D. M. Pardoll, "Eos mediates Foxp3-dependent gene silencing in CD4+ regulatory T cells," *Science*, vol. 325, no. 5944, pp. 1142–6, 2009.

[75] J. R. DiSpirito and H. Shen, "Expression analysis of resting and stimulated naïve and MP CD8 T cells." `https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE16145`, 2012.

[76] X. Liu, X. Qu, Y. Chen, L. Liao, K. Cheng, C. Shao, M. Zenke, A. Keating, and R. C. Zhao, "Mesenchymal stem/stromal cells induce the generation of novel IL-10-dependent regulatory dendritic cells by SOCS3 activation," *J Immunol*, vol. 189, no. 3, pp. 1182–92, 2012.

[77] K. Al Moussawi, E. Ghigo, U. Kalinke, L. Alexopoulou, J. L. Mege, and B. Desnues, "Type I interferon induction is detrimental during infection with the Whipple's disease bacterium, Tropheryma whipplei," *PLoS Pathog*, vol. 6, no. 1, p. e1000722, 2010.

[78] E. Ghigo, A. O. Barry, L. Pretat, K. Al Moussawi, B. Desnues, C. Capo, H. Kornfeld, and J. L. Mege, "IL-16 promotes T. whipplei replication by inhibiting phagosome conversion and modulating macrophage activation," *PLoS One*, vol. 5, no. 10, p. e13561, 2010.

[79] F. K. Swirski, M. Nahrendorf, M. Etzrodt, M. Wildgruber, V. Cortez-Retamozo, P. Panizzi, J. L. Figueiredo, R. H. Kohler, A. Chudnovskiy, P. Waterman, E. Aikawa, T. R. Mempel, P. Libby, R. Weissleder, and M. J. Pittet, "Identification of splenic reservoir monocytes and their deployment to inflammatory sites," *Science*, vol. 325, no. 5940, pp. 612–6, 2009.

[80] A. O. Latorre, B. D. Caniceiro, H. Fukumasu, D. R. Gardner, F. M. Lopes, J. Wysochi, H. L., T. C. da Silva, M. Haraguchi, F. F. Bressan, and S. L. Gorniak, "Ptaquiloside reduces NK cell activities by enhancing metallothionein expression, which is prevented by selenium," *Toxicology*, vol. 304, pp. 100–8, 2013.

[81] T. S. Heng, M. W. Painter, and C. Immunological Genome Project, "The Immunological Genome Project: networks of gene expression in immune cells," *Nat Immunol*, vol. 9, no. 10, pp. 1091–4, 2008.

[82] P. Langfelder and S. Horvath, "WGCNA: an R package for weighted correlation network analysis," *BMC Bioinformatics*, vol. 9, p. 559, 2008.

[83] K. Hasegawa, A. Sato, K. Tanimura, K. Uemasu, Y. Hamakawa, Y. Fuseya, S. Sato, S. Muro, and T. Hirai, "Fraction of MHCII and EpCAM expression characterizes distal lung epithelial cells for alveolar type 2 cell isolation," *Respir Res*, vol. 18, no. 1, p. 150, 2017.

[84] T. Guirimand, S. Delmotte, and V. Navratil, "VirHostNet 2.0: surfing on the web of virus/host molecular interactions data," *Nucleic Acids Res*, vol. 43, no. Database issue, pp. D583–7, 2015.

[85] A. Duran, J. F. Linares, A. S. Galvez, K. Wikenheiser, J. M. Flores, M. T. Diaz-Meco, and J. Moscat, "The signaling adaptor p62 is an important NF-kappaB mediator in tumorigenesis," *Cancer Cell*, vol. 13, no. 4, pp. 343–54, 2008.

[86] J. M. Scheffler, F. Sparber, C. H. Tripp, C. Herrmann, A. Humenberger, J. Blitz, N. Romani, P. Stoitzner, and L. A. Huber, "LAMTOR2 regulates dendritic cell homeostasis through FLT3-dependent mTOR signalling," *Nat Commun*, vol. 5, p. 5138, 2014.

[87] W. Zhang and H. T. Liu, "MAPK signal pathways in the regulation of cell proliferation in mammalian cells," *Cell Res*, vol. 12, no. 1, pp. 9–18, 2002.

[88] M. Muller, S. Carter, M. J. Hofer, and I. L. Campbell, "Review: The chemokine receptor CXCR3 and its ligands CXCL9, CXCL10 and CXCL11 in neuroimmunity–a tale of conflict and conundrum," *Neuropathol Appl Neurobiol*, vol. 36, no. 5, pp. 368–87, 2010.

[89] M. Kiso, T. J. S. Lopes, S. Yamayoshi, M. Ito, M. Yamashita, N. Nakajima, H. Hasegawa, G. Neumann, and Y. Kawaoka, "Combination Therapy With Neuraminidase and Polymerase Inhibitors in Nude Mice Infected With Influenza Virus," *Journal of Infectious Diseases*, vol. 217, no. 6, pp. 887–896, 2018.

[90] H. Ueki, I. H. Wang, S. Fukuyama, H. Katsura, T. J. da Silva Lopes, G. Neumann, and Y. Kawaoka, "In vivo imaging of the pathophysiological changes and neutrophil dynamics in influenza virus-infected mouse lungs," *Proc Natl Acad Sci U S A*, vol. 115, no. 28, pp. E6622–E6629, 2018.

[91] K. Iwatsuki-Horimoto, N. Nakajima, Y. Ichiko, Y. Sakai-Tagawa, T. Noda, H. Hasegawa, and Y. Kawaoka, "Syrian Hamster as an Animal Model for the Study of Human Influenza Virus Infection," *Journal of Virology*, vol. 92, no. 4, 2018.

[92] J. A. Briggs, C. Weinreb, D. E. Wagner, S. Megason, L. Peshkin, M. W. Kirschner, and A. M. Klein, "The dynamics of gene expression in vertebrate embryogenesis at single-cell resolution," *Science*, vol. 360, no. 6392, 2018.

[93] A. Schuhmacher, O. Gassmann, and M. Hinder, "Changing R&D models in research-based pharmaceutical companies," *J Transl Med*, vol. 14, no. 1, p. 105, 2016.

[94] S. Pushpakom, F. Iorio, P. A. Eyers, K. J. Escott, S. Hopper, A. Wells, A. Doig, T. Guilliams, J. Latimer, C. McNamee, A. Norris, P. Sanseau, D. Cavalla, and M. Pirmohamed, "Drug repurposing: progress, challenges and recommendations," *Nat Rev Drug Discov*, vol. 18, no. 1, pp. 41–58, 2019.

[95] S. M. Paul, D. S. Mytelka, C. T. Dunwiddie, C. C. Persinger, B. H. Munos, S. R. Lindborg, and A. L. Schacht, "How to improve R&D productivity: the pharmaceutical industry's grand challenge," *Nat Rev Drug Discov*, vol. 9, no. 3, pp. 203–14, 2010.

[96] J. A. DiMasi, H. G. Grabowski, and R. W. Hansen, "Innovation in the pharmaceutical industry: New estimates of R&D costs," *J Health Econ*, vol. 47, pp. 20–33, 2016.

[97] M. Hay, D. W. Thomas, J. L. Craighead, C. Economides, and J. Rosenthal, "Clinical development success rates for investigational drugs," *Nat Biotechnol*, vol. 32, no. 1, pp. 40–51, 2014.

[98] T. J. Hwang, D. Carpenter, J. C. Lauffenburger, B. Wang, J. M. Franklin, and A. S. Kesselheim, "Failure of Investigational Drugs in Late-Stage Clinical Development and Publication of Trial Results," *JAMA Intern Med*, vol. 176, no. 12, pp. 1826–1833, 2016.

[99] P. Spagnolo and T. M. Maher, "Clinical trial research in focus: why do so many clinical trials fail in IPF?," *Lancet Respir Med*, vol. 5, no. 5, pp. 372–374, 2017.

[100] H. Naci and J. P. Ioannidis, "How good is "evidence" from clinical studies of drug effects and why might such evidence fail in the prediction of the clinical utility of drugs?," *Annu Rev Pharmacol Toxicol*, vol. 55, pp. 169–89, 2015.

[101] I. Gashaw, P. Ellinghaus, A. Sommer, and K. Asadullah, "What makes a good drug target?," *Drug Discovery Today*, vol. 17, pp. S24–S30, 2012.

[102] H. N. Chua and F. P. Roth, "Discovering the targets of drugs via computational systems biology," *J Biol Chem*, vol. 286, no. 27, pp. 23653–8, 2011.

[103] J. Lamb, E. D. Crawford, D. Peck, J. W. Modell, I. C. Blat, M. J. Wrobel, J. Lerner, J. P. Brunet, A. Subramanian, K. N. Ross, M. Reich, H. Hieronymus, G. Wei, S. A. Armstrong, S. J. Haggarty, P. A. Clemons, R. Wei, S. A. Carr, E. S. Lander, and T. R. Golub, "The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease," *Science*, vol. 313, no. 5795, pp. 1929–35, 2006.

[104] B. Ganter, S. Tugendreich, C. I. Pearson, E. Ayanoglu, S. Baumhueter, K. A. Bostian, L. Brady, L. J. Browne, J. T. Calvin, G. J. Day, N. Breckenridge, S. Dunlea, B. P. Eynon, L. M. Furness, J. Ferng, M. R. Fielden, S. Y. Fujimoto, L. Gong, C. Hu, R. Idury, M. S. Judo, K. L. Kolaja, M. D. Lee, C. McSorley, J. M. Minor, R. V. Nair, G. Natsoulis, P. Nguyen, S. M. Nicholson, H. Pham, A. H. Roter, D. Sun, S. Tan, S. Thode, A. M. Tolley, A. Vladimirova, J. Yang, Z. Zhou, and K. Jarnagin, "Development of a large-scale chemogenomics database to improve drug candidate selection and to understand mechanisms of chemical toxicity and action," *J Biotechnol*, vol. 119, no. 3, pp. 219–44, 2005.

[105] M. Wang, C. Tang, and J. Chen, "Drug-Target Interaction Prediction via Dual Laplacian Graph Regularized Matrix Completion," *Biomed Res Int*, vol. 2018, p. 1425608, 2018.

[106] A. Vertes, A.-B. Arul, P. Avar, A. R. Korte, H. Li, P. Nemes, L. Parvin, S. Stopka, S. Hwang, Z. J. Sahab, L. Zhang, D. I. Bunin, M. Knapp, A. Poggio, M.-O. Stehr, C. L. Talcott, B. M. Davis, S. R. Dinn, C. A. Morton, C. J. Sevinsky, and M. I. Zavodszky, "Inferring Mechanism of Action of an Unknown Compound from Time Series Omics Data," Computational Methods in Systems Biology, pp. 238–255, Springer International Publishing.

[107] A. J. Wolpaw, K. Shimada, R. Skouta, M. E. Welsch, U. D. Akavia, D. Pe'er, F. Shaik, J. C. Bulinski, and B. R. Stockwell, "Modulatory profiling identifies mechanisms of small molecule-induced cell death," *Proc Natl Acad Sci U S A*, vol. 108, no. 39, pp. E771–80, 2011.

[108] F. Iorio, T. Rittman, H. Ge, M. Menden, and J. Saez-Rodriguez, "Transcriptional data: a new gateway to drug repositioning?," *Drug Discov Today*, vol. 18, no. 7-8, pp. 350–7, 2013.

[109] S. T. A. Rush and D. Repsilber, "Capturing context-specific regulation in molecular interaction networks," *BMC Bioinformatics*, vol. 19, no. 1, p. 539, 2018.

[110] E. Y. Chen, H. Xu, S. Gordonov, M. P. Lim, M. H. Perkins, and A. Ma'ayan, "Expression2Kinases: mRNA profiling linked to multiple upstream regulatory layers," *Bioinformatics*, vol. 28, no. 1, pp. 105–11, 2012.

[111] M. Koido, Y. Tani, S. Tsukahara, Y. Okamoto, and A. Tomida, "InDePTH: detection of hub genes for developing gene expression networks under anticancer drug treatment," *Oncotarget*, vol. 9, no. 49, pp. 29097–29111, 2018.

[112] X. Ji, J. M. Freudenberg, and P. Agarwal, "Integrating Biological Networks for Drug Target Prediction and Prioritization," *Methods Mol Biol*, vol. 1903, pp. 203–218, 2019.

[113] E. J. Cosgrove, Y. Zhou, T. S. Gardner, and E. D. Kolaczyk, "Predicting gene targets of perturbations via network-based filtering of mRNA expression compendia," *Bioinformatics*, vol. 24, no. 21, pp. 2482–90, 2008.

[114] M. Failli, J. Paananen, and V. Fortino, "Prioritizing target-disease associations with novel safety and efficacy scoring methods," *Sci Rep*, vol. 9, no. 1, p. 9852, 2019.

[115] D. Szklarczyk, A. L. Gable, D. Lyon, A. Junge, S. Wyder, J. Huerta-Cepas, M. Simonovic, N. T. Doncheva, J. H. Morris, P. Bork, L. J. Jensen, and C. V. Mering,

"STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets," *Nucleic Acids Res*, vol. 47, no. D1, pp. D607–D613, 2019.

[116] P. Cahan, H. Li, S. A. Morris, E. Lummertz da Rocha, G. Q. Daley, and J. J. Collins, "CellNet: network biology applied to stem cell engineering," *Cell*, vol. 158, no. 4, pp. 903–915, 2014.

[117] M. Zhu, L. Gao, X. Li, Z. Liu, C. Xu, Y. Yan, E. Walker, W. Jiang, B. Su, X. Chen, and H. Lin, "The analysis of the drug–targets based on the topological properties in the human protein–protein interaction network," *Journal of Drug Targeting*, vol. 17, no. 7, pp. 524–532, 2009.

[118] T. J. Lopes, J. E. Shoemaker, Y. Matsuoka, Y. Kawaoka, and H. Kitano, "Identifying problematic drugs based on the characteristics of their targets," *Front Pharmacol*, vol. 6, p. 186, 2015.

[119] E. E. Ackerman, E. Kawakami, M. Katoh, T. Watanabe, S. Watanabe, Y. Tomita, T. J. Lopes, Y. Matsuoka, H. Kitano, J. E. Shoemaker, and Y. Kawaoka, "Network-Guided Discovery of Influenza Virus Replication Host Factors," *MBio*, vol. 9, no. 6, 2018.

[120] Y. Feng, Q. Wang, and T. Wang, "Drug Target Protein-Protein Interaction Networks: A Systematic Perspective," *Biomed Res Int*, vol. 2017, p. 1289259, 2017.

[121] M. Bansal, J. Yang, C. Karan, M. P. Menden, J. C. Costello, H. Tang, G. Xiao, Y. Li, J. Allen, R. Zhong, B. Chen, M. Kim, T. Wang, L. M. Heiser, R. Realubit, M. Mattioli, M. J. Alvarez, Y. Shen, N.-D. Community, D. Gallahan, D. Singer, J. Saez-Rodriguez, Y. Xie, G. Stolovitzky, A. Califano, and N.-D. Community, "A community computational challenge to predict the activity of pairs of compounds," *Nat Biotechnol*, vol. 32, no. 12, pp. 1213–22, 2014.

[122] C. Magkoufopoulou, S. M. Claessen, M. Tsamou, D. G. Jennen, J. C. Kleinjans, and J. H. van Delft, "A transcriptomics-based in vitro assay for predicting chemical genotoxicity in vivo," *Carcinogenesis*, vol. 33, no. 7, pp. 1421–9, 2012.

[123] S. Kubicek, J. C. Gilbert, D. Fomina-Yadlin, A. D. Gitlin, Y. Yuan, F. F. Wagner, E. B. Holson, T. Luo, T. A. Lewis, B. Taylor, S. Gupta, A. F. Shamji, B. K. Wagner, P. A. Clemons, and S. L. Schreiber, "Chromatin-targeting small molecules cause class-specific transcriptional changes in pancreatic endocrine cells," *Proc Natl Acad Sci U S A*, vol. 109, no. 14, pp. 5364–9, 2012.

[124] L. Gautier, L. Cope, B. M. Bolstad, and R. A. Irizarry, "affy—analysis of Affymetrix GeneChip data at the probe level," *Bioinformatics*, vol. 20, no. 3, pp. 307–315, 2004.

[125] M. Carlson, *hgu219.db: Affymetrix Human Genome 219 Plate annotation data (chip hgu219)*, 2016. R package version 3.2.3.

[126] M. Carlson, *moe430a.db: Affymetrix Mouse Expression Set 430 annotation data (chip moe430a)*, 2016. R package version 3.2.3.

[127] D. Marbach, D. Lamparter, G. Quon, M. Kellis, Z. Kutalik, and S. Bergmann, "Tissue-specific regulatory circuits reveal variable modular perturbations across complex diseases," *Nat Methods*, vol. 13, no. 4, pp. 366–70, 2016.

[128] H. Han, J. W. Cho, S. Lee, A. Yun, H. Kim, D. Bae, S. Yang, C. Y. Kim, M. Lee, E. Kim, S. Lee, B. Kang, D. Jeong, Y. Kim, H. N. Jeon, H. Jung, S. Nam, M. Chung, J. H. Kim, and I. Lee, "TRRUST v2: an expanded reference database of human and mouse transcriptional regulatory interactions," *Nucleic Acids Res*, vol. 46, no. D1, pp. D380–D386, 2018.

[129] Z. P. Liu, C. Wu, H. Miao, and H. Wu, "RegNetwork: an integrated database of transcriptional and post-transcriptional regulatory networks in human and mouse," *Database (Oxford)*, vol. 2015, 2015.

[130] G. Csardi and T. Nepusz, "The Igraph Software Package for Complex Network Research," *InterJournal*, vol. Complex Systems, p. 1695, 2005.

[131] M. Kuhn, C. von Mering, M. Campillos, L. J. Jensen, and P. Bork, "STITCH: interaction networks of chemicals and proteins," *Nucleic acids research*, vol. 36, no. Database issue, pp. D684–D688, 2008.

[132] D. Szklarczyk, A. Santos, C. von Mering, L. J. Jensen, P. Bork, and M. Kuhn, "STITCH 5: augmenting protein-chemical interaction networks with tissue and affinity data," *Nucleic acids research*, vol. 44, no. D1, pp. D380–D384, 2016.

[133] X. Robin, N. Turck, A. Hainard, N. Tiberti, F. Lisacek, J.-C. Sanchez, and M. Müller, "pROC: an open-source package for R and S+ to analyze and compare ROC curves," *BMC Bioinformatics*, vol. 12, no. 1, p. 77, 2011.

[134] J. Ivanic, X. Yu, A. Wallqvist, and J. Reifman, "Influence of protein abundance on high-throughput protein-protein interaction detection," *PLoS One*, vol. 4, no. 6, p. e5815, 2009.

[135] T. G. O. Consortium, "The Gene Ontology Resource: 20 years and still GOing strong," *Nucleic Acids Res*, vol. 47, no. D1, pp. D330–D338, 2019.

[136] M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin, and G. Sherlock, "Gene ontology: tool for the unification of biology," *Nat Genet*, vol. 25, no. 1, pp. 25–9, 2000.